

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 22, 2003, 20:22:25 ; Search time 104 Seconds

Sequence: (without alignments) 1347.332 Million cell updates/sec

Title: US-09-759-207-2

Perfect score: 2142

Sequence: MILRSKPALPPPLMLLGP.....LPAFSYSFFVIRNAKVAACI 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRIMBL 23.0

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_invertebrate:\*
- 5: sp\_mammal:\*
- 6: sp\_mhc:\*
- 7: sp\_organelle:\*
- 8: sp\_phage:\*
- 9: sp\_plant:\*
- 10: sp\_rhod:\*
- 11: sp\_rabbit:\*
- 12: sp\_virus:\*
- 13: sp\_verebrale:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Score Query Length DB ID Description

Result No.	Score	Query	Length	DB	ID	Description
1	2838	99.9	543	4	09Y251	09Y251 homo sapien
2	2817	99.1	545	4	09UL39	09UL39 homo sapien
3	2250	80.3	545	6	09MY0	09MY0 bov taurus
4	2123	75.7	535	11	09QZK8	09QZK8 mus musculus
5	1645.5	74.7	536	11	09QZK8	09QZK8 rattus norvegicus
6	1154.5	57.9	523	13	09QYK5	09QYK5 gallus gallus
7	1146.5	40.6	592	4	09HB37	09HB37 homo sapien
8	1015.5	40.3	592	4	08WQ02	08WQ02 homo sapien
9	936.5	35.7	548	4	08WWQ1	08WWQ1 homo sapien
10	897.5	31.6	480	4	09HB38	09HB38 homo sapien
11	696	24.5	515	5	08T108	08T108 bombyx mori
12	416	14.6	543	10	09SDA1	09SDA1 arabidopsis
13	392.5	13.8	544	10	09FP10	09FP10 arabidopsis
14	381	13.4	527	10	09LR15	09LR15 oryza sativa
15					09LR18	09LR18 scutellaria

Database :

SPTRIMBL 23.0

PRELIMINARY

ID 09Y251 PRT: 543 AA.

AC 09Y251;

DT 01-NOV-1999 (TREMBBLE 12, Created)

DT 01-OCT-2002 (TREMBBLE 22, Last annotation update)

DE HEPARANASE.

GN HPA.

OS Homo sapiens (Human).

OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

[1] SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE:9321249; PubMed:10395326;

RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

RA Parish C.R., Cloning of mammalian heparanase, an important enzyme in tumor

RT invasion and metastasis., Nat. Med. 5:803-809(1999).

RN [2] SEQUENCE FROM N.A.

RP Vlodavsky I., Friedman Y., Elkin M., Ainsztein H., Atzmon R.,

RA Ishai-Michaeli R., Bitan M., Rapaport O., Peretz T., Michal I.,

RA Spector I., Becker I., "Mammalian heparanase: a novel gene involved in tumor progression and

RA metastasis.", Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.

RN [3] SEQUENCE FROM N.A. MEDLINE:9377052; PubMed:10446189;

RA Toyoshima M., Nakajima M., "Human heparanase. Purification, characterization, cloning, and

RA expression.", J. Biol. Chem. 274:24153-24160(1999).

RN [4] SEQUENCE FROM N.A.

17	363	12.8	-536	10	09FP11	09FP11 arabidopsis
18	352.5	12.4	-516	10	09FLKB	09FLKB arabidopsis
19	352.5	12.4	-539	10	08L160	08L160 arabidopsis
20	169.5	6.0	-190	10	082604	082604 arabidopsis
21	160	5.6	-935	5	09VE79	09ve79 drosophila
22	130.5	4.6	-493	17	09HK01	09hk01 thermoplasm
23	122.5	4.3	-408	3	09HEZ1	09hez1 phanerochaet
24	122.5	4.3	-408	3	09HEZ2	09hez2 phanerochaet
25	116.5	4.1	-617	12	040996	040996 measles vir
26	114	4.0	-398	16	08F410	08F410 leptospira
27	113.5	4.0	-617	12	083295	083295 measles vir
28	113.5	4.0	-1829	2	09KH44	09kh44 parvovirus
29	113	4.0	-390	17	08TPH7	08tpm7 methanococcus
30	112.5	4.0	-617	12	083647	083647 neurospora
31	111.5	3.9	-493	17	097907	097907 encephalitozoon
32	111.5	3.9	-575	10	0911P6	0911P6 measles vir
33	111.5	3.9	-617	12	0910N9	0910N9 measles vir
34	111.5	3.9	-617	12	098VH5	098vh5 measles vir
35	111	3.9	-670	10	09M090	09m090 arabidopsis
36	111	3.9	-2319	3	096U00	096u00 neurospora
37	110.5	3.9	-475	17	088U17	088u17 encephalitozoon
38	109.5	3.9	-617	12	0911P6	0911P6 measles vir
39	109.5	3.9	-617	12	0910N9	0910N9 measles vir
40	105.5	3.9	-879	16	09XCP4	09xcp4 escherichia
41	109	3.8	-617	12	098VU6	098vu6 measles vir
42	108.5	3.8	-411	16	072895	072895 synechocystis
43	108.5	3.8	-617	12	011381	011381 measles vir
44	107.5	3.8	-500	16	09A5U0	09a5u0 caulobacter
45	107.5	3.8	-617	12	09Q93	09q93 measles vir

## ALIGNMENTS

RC	SEQUENCE FROM N.A.
RE	TISSUE=Placenta;
RE	REDBLNE=9935379; PubMed=040543;
RA	Kuskie, P.H.; Hulmes, J.D.; Ludwig, D.; Patel, S.; Navarro, E.C.; Seddon, A.P.; Giorgio, N.A.; Bonien, P.; "Cloning and Functional Expression of a Human Heparanase Gene," Biomed. Biophys. Res. Commun., 261:183-187(1999).
RL	DR EMBL; AF16154; ADD45379; 1; DR EMBL; AF14325; ADD43421; 1; DR EMBL; AF15510; ADD49411; 1; DR EMBL; AF15376; ADD45669; 1; DR Genew; HGNC:5164; HRB; DR InterPro; IPR05199; Glyco_hydro_79n.
DR	PFam; PF03662; Glyco_hydro_79n; 1; DR
DR	SEQUENCE 543 AA; 61176 MW; AD262BC267334AB2 CRC64;
Query Match	99.9%; Score 2839; DB 4; Length 543;
Best Local Similarity	99.8%;
Matches	542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MLLRSKPALPPMLLUGPCLPSPGALPRAQAOQDVDFPFTQBLHVSPLSPSUT 60
Db	1 MLLRSKPALPPMLLUGPCLPSPGALPRAQAOQDVDFPFTQBLHVSPLSPSUT 60
Qy	61 IDANLATAPRELILASPKLRTLARGSPAYLRFGGTKTDFLFDPKKESTEERSWQS 120
Db	61 IDANLATAPRELILASPKLRTLARGSPAYLRFGGTKTDFLFDPKKESTEERSWQS 120
Qy	121 QWNQDICKYGSIPPDVBEKRLLEWPKQEQOLLREHBYQKFKNSTYSSVDSLTYFAN 180
Db	121 QWNQDICKYGSIPPDVBEKRLLEWPKQEQOLLREHBYQKFKNSTYSSVDSLTYFAN 180
Qy	181 GIDLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFNS 240
Db	181 GIDLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFNS 240
Qy	241 QLGEDYIOLHKLURKSTPKNAKLYGPDVGQPRRTAKMLKSLFKAGGEVIDSVTHWY 300
Db	241 QLGEDYIOLHKLURKSTPKNAKLYGPDVGQPRRTAKMLKSLFKAGGEVIDSVTHWY 300
Qy	301 NGRTATREDFLNPDVDFISSIONVQKQFQVQVESTRPGKWWLGETSSAYGGAPLSDT 360
Db	301 NGRTATREDFLNPDVDFISSIONVQKQFQVQVESTRPGKWWLGETSSAYGGAPLSDT 360
Qy	361 AGFMWWDKLGISARMGTEVNMQFREGAGNYILVEDENFDPLDWDYSLFLKGUVGTVL 420
Db	361 AGFMWWDKLGISARMGTEVNMQFREGAGNYILVEDENFDPLDWDYSLFLKGUVGTVL 420
Qy	421 ASVOGSKRKURKVLYLCTNTDNPYKEGDLTYAINLHNVTKLRLPPSNQKVQDVKLL 480
Db	421 ASVOGSKRKURKVLYLCTNTDNPYKEGDLTYAINLHNVTKLRLPPSNQKVQDVKLL 480
Qy	481 RPLPGPGGLSKSVOLNGLTKRAVDDQTLPLMEKPLRGSSIGLPAFSSYFFVIRNAKA 540
Db	481 RPLPGPGGLSKSVOLNGLTKRAVDDQTLPLMEKPLRGSSIGLPAFSSYFFVIRNAKA 540
Qy	541 ACI 543
Db	541 ACI 543
Db	541 ACI 543
RESULT 2	RESULT 3
Q9UNJ9	Q9UNJ9
PRIMINARY;	PRIMINARY;
PRT;	PRT;
545 AA.	545 AA.
ID	Q9UNJ9
AC	Q9UNJ9;
DT	01-MAY-2000 (TREMBrel. 13, Created)
DT	01-MAY-2000 (TREMBrel. 13, Last sequence update)
DT	01-JUN-2002 (TREMBrel. 21, Last annotation update)
DE	Heparanase.
OS	Bos taurus (Bovine).
OC	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi; Homo sapiens (Human).
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TAXID=9606;	
[1]	
RC	SEQUENCE FROM N.A.:
RC	TISSUE=Placenta;
RC	REDBLNE=2029546; Pubmed=10764835;
RA	Dempsey L.A.; Plummer T.B.; Combes S.L.; Platt J.L.; "Reparanase expression in invasive trophoblasts and acute vascular damage," Glycobiology 10:467-475(2000).
RL	DR EMBL; AF084467; ADD54316; 1; DR InterPro; IPR05199; Glyco_hydro_79n.
DR	DR PFam; PF03662; Glyco_hydro_79n; 1; DR
DR	SEQUENCE 545 AA; 61418 MW; 67B80ACD73CSA9A1 CRC64;
Query Match	99.1%; Score 2817; DB 4; Length 545;
Best Local Similarity	99.4%;
Matches	542; Conservative 1; Mismatches 0; Indels 2; Gaps 1;
Qy	1 MLLRSKPALPPMLLUGPCLPSPGALPRAQAOQDVDFPFTQBLHVSPLSPSUT 60
Db	1 MLLRSKPALPPMLLUGPCLPSPGALPRAQAOQDVDFPFTQBLHVSPLSPSUT 60
Qy	59 VTDANLATAPRELILASPKLRTLARGSPAYLRFGGTKTDFLFDPKKESTEERSW 110
Db	61 VTDANLATAPRELILASPKLRTLARGSPAYLRFGGTKTDFLFDPKKESTEERSW 110
Qy	119 QSQNQDICKYGSIPPDVBEKRLLEWPKQEQOLLREHBYQKFKNSTYSSVDSLTYFAN 170
Db	121 QSQNQDICKYGSIPPDVBEKRLLEWPKQEQOLLREHBYQKFKNSTYSSVDSLTYFAN 170
Qy	179 CSGDLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFIN 230
Db	181 CSGDLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFIN 230
Qy	239 GSGDGLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFIN 290
Db	241 GSGDGLDRGFLNALLRTADLOWNSNAQQLDYCSSKGYNISWEIGNEPNPLKGADIFIN 290
Qy	299 YNGRTATREDFLNPDVDFISSIONVQKQFQVQVESTRPGKWWLGETSSAYGGAPLSDT 350
Db	301 YNGRTATREDFLNPDVDFISSIONVQKQFQVQVESTRPGKWWLGETSSAYGGAPLSDT 350
Qy	359 FAAGFMWWDKLGISARMGTEVNMQFREGAGNYILVEDENFDPLDWDYSLFLKGUVGTV 410
Db	361 FAAGFMWWDKLGISARMGTEVNMQFREGAGNYILVEDENFDPLDWDYSLFLKGUVGTV 410
Qy	419 LMASVQGSKRKURKVLYLCTNTDNPYKEGDLTYAINLHNVTKLRLPPSNQKVQDVKY 470
Db	421 LMASVQGSKRKURKVLYLCTNTDNPYKEGDLTYAINLHNVTKLRLPPSNQKVQDVKY 470
Qy	479 LRLPLGPGGLSKSVOLNGLTKRAVDDQTLPLMEKPLRGSSIGLPAFSSYFFVIRNAK 530
Db	481 LRLPLGPGGLSKSVOLNGLTKRAVDDQTLPLMEKPLRGSSIGLPAFSSYFFVIRNAK 530
Qy	539 VAACI 543
Db	541 VAACI 543

RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;  
 RA "Expression of Heparanase mRNA in Bovine Placenta During Gestation."  
 RT Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 RL EMBL:AP281160; AAF87301; 2;  
 DR Intrpct: IPR005199; Glyco\_hydro\_79N.  
 PRM: PF03622; Glyco\_hydro\_79n; I.  
 SEQUENCE 545-AA; 61076 MW; FAC4BDFD855B933 CRC64;

Query Match 80.3%; Score 2282; DB 6; Length 545;  
 Best Local Similarity 79.8%; Pred. No. 9; 5e-174; Mismatches 73; Indels 2; Gaps 1;  
 Matches 435; Conservative 35; Mismatches 73; Indels 2; Gaps 1;

QY 1 MURSKRKLPPPMML-LIGPGPGSPGALPRAQADQVVDLDPFTQRPPLHVSPLS 58  
 Db 1 MURCKEGLRPPPLLPLPGLPGCPGTPAAAPADABELEBFPTRPLHVSPLS 60

QY 59 VTPDANLATPPELILGSPKURTLGLSPVPLREGGTTKDPKESTERSYW 118  
 Db 61 PRIDANLATDPPELILGSPKURTLGLSPVPLREGGTTKDPKESTERSYW 120

QY 119 QSQVNQDICKYSSIPDVEKURLEWPIQEQQLRHEHQKEKNTYSRSSVUDVLYTAN 178  
 Db 121 LSQSNQDICKSSIPSQVEKURLEWPIQEQQLRHEHQKEKNTYSRSSVUDVLYTAN 180

QY 179 CSGDLIPLGQNLMLRITADLOMNSNAQOLLYDCCSKGKNTYSWELGNEPNPLKCADIFIN 238  
 Db 181 CSGQNLIFGQNLMLRITDHWQWSNAQOLLYDCCSKGKNTYSWELGNEPNPLKCADIFIN 240

QY 239 GSQLGRLYIQQLKLURSTPKQAKLYPDPGCPERKAKOMKSFKAQGEVDSVWPHY 298  
 Db 241 GRCQGEDETEFRILGKSAFQNLKLYSPDQCPERKAKOMKSFKAQGEVDSVWPHY 300

QY 299 YLNGRTATREDFNPDVUDIFSSVQKVQFVSTRCKWKGEGTSSAYGGAPLSDT 358  
 Db 301 YNGRATKEDFNLDPDFTFSQVKLRLTKEKIRPKWKGEGTSSAFCGGAPLSDT 360

QY 359 RAQGMWMLDKLGSARMGIEVNRQVFFAGGNHLVUDENFDPLDYMISLPLFKLVGTV 418  
 Db 361 FAGFWMWMLDKLGSARMGIEVNRQVFFAGGNHLVUDENFDPLDYMISLPLFKLVGTV 420

QY 419 LMSVQGSKRKRRKVLHCTNTDNPYKEGDTLYAHLNHTVTKLRLPYPENPKDVKY 478  
 Db 421 LMSVQGSKRKRRKVLHCTNTDNPYKEGDTLYAHLNHTVTKLRLPYPENPKDVKY 480

QY 479 LIRPLGPGLUSKSVQVNGLTLKQVDDOTLPPMEKPKRGSSGLUPAFSSFFVIRNAK 538  
 Db 481 LIRPSGTLGUSKSVQVNGLTLKQVDDOTLPPMEKPKRGSSGLUPAFSSFFVIRNAK 540

QY 539 VACI 543  
 Db 541 VACI 545

RESULT 4

Q8K3K3 PRELIMINARY; PRT; 535 AA.

ID Q8K3K3  
 AC Q8K3K3;  
 DT 01-OCT-2002 (TREMBrel. 22, Created)  
 DT 01-OCT-2002 (TREMBrel. 22, Last sequence update)  
 DB Heparanase.  
 GN HBB OR HPA.  
 OS Mus musculus (Mouse).  
 OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Ruthereria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1];  
 RP SOURCECB FROM N.A.  
 RC STRAIN=FBV;  
 RA Mao H.-O., Navarro B., Patel S., Sargent D., Koo H., Wan H., Plata A., Zhou O., Ludwig D., Boilen P., Kusse P.;  
 RT Cloning, expression, and purification of mouse heparanase.;  
 RL Protein Expr. Purif. 0:0-0(2002).

RN [2];  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=2235483; PubMed=12466851;  
 RA The FANTOM Consortium  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002);  
 DR EMBL; AY077467; AAU76083; 1;  
 DR EMBL; AK004071; BAC30600; 1;  
 DR MGI; MGI:1343124; Bipe.  
 DR InterPro; IPR005199; Glyco\_hydro\_79n; I.  
 DR Pfam; PF03622; Glyco\_hydro\_79n; I.  
 SQ SEQUENCE 535 AA; 60655 MW; 6873A8302FB8A0DF CRC64;

Query Match 75.7%; Score 2190; DB 11; Length 535;  
 Best Local Similarity 76.6%; Pred. No. 3; 3e-163; Mismatches 73; Indels 0; Gaps 0; Matches 407; Conservative 51; Mismatches 73; Indels 0; Gaps 0;

QY 13 LMLLIGPIGLSPGALPRAQADQVVDLDPFTQRPPLHVSPLSVDANLATPREF 72  
 Db 135 LLLWLGPIGLSPGALPRAQADQVVDLDPFTQRPPLHVSPLSVDANLATPREF 64

QY 173 ILGSISKLRLTARGSPAYRFGGKTDIFLDPKESTERSYWQSQVNQDICKYGI 132  
 Db 175 TFLGSPRIRALRGLSPAYRFGGKTDIFLDPKESTERSYWQSQVNQDICKYGI 124

QY 193 RTADLOMNSNAQOLLYDCCSKGKNTYSWELGNEPNPLKCADIFINGSGEDYOLHKU 252  
 Db 195 SAAVLRLQVNPFOBLRLLREQYKEFKFYSTYSRSSVUDVLYTANHNTVTKLRLPYPENPKDVKY 244

QY 253 LRKSTKRNALKYGPDVQPRKAKMKTSLKAGEVDSVTHHYLNGRTATREDFIN 312  
 Db 245 LQSAFQNLKLYGPDVQPRKAKMKTSLKAGEVDSVTHHYLNGRTATREDFIN 304

QY 313 PDVLDIFISVQVYVWESVTPRGKWMQRTSSAYGGAPLSDTFAANGFMWLDKLGIS 372  
 Db 315 SDVLDIFISVQVYVWESVTPRGKWMQRTSSAYGGAPLSDTFAANGFMWLDKLGIS 364

QY 373 ARMGIEVUMRQVFFAGGNHLVUDENFDPLDYMISLPLFKLVGTVLMSVQSGSKRKLR 432  
 Db 375 VYHCTVYHRYQEGDULYVNLNHTVTKLRLPPLKPKVPLKSGPGLSKS 484

QY 433 VYHCTVYHRYQEGDULYVNLNHTVTKLRLPPLKPKVPLKSGPGLSKS 492  
 Db 425 VYHCTVYHRYQEGDULYVNLNHTVTKLRLPPLKPKVPLKSGPGLSKS 484

QY 493 VYHCTVYHRYQEGDULYVNLNHTVTKLRLPPLKPKVPLKSGPGLSKS 543  
 Db 485 VYHCTVYHRYQEGDULYVNLNHTVTKLRLPPLKPKVPLKSGPGLSKS 535

RESULT 5

Q9QZFB PRELIMINARY; PRT; 536 AA.

ID Q9QZFB  
 AC Q9QZFB;  
 DT 01-MAY-2000 (TREMBrel. 13, Created)  
 DT 01-MAY-2000 (TREMBrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBrel. 13, Last annotation update)  
 DT 01-JUN-2002 (TREMBrel. 21, Last annotation update)  
 DB Heparanase.  
 GN HEP.  
 OS Rattus norvegicus (Rat).  
 OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Ruthereria; Rodentia; Sciuromorphi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1];  
 RP SEQUENCE FROM N.A.  
 RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;



RESULT 8

Q8WQ2	PRELIMINARY;	PRT;	592 AA.
AC	Q8WQ2;		
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2003 (Tremblrel. 20, Last sequence update)		
DB	Heparase 2.		
OS	Homo sapiens (Human)		
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RC	Pessegue Safontas B.J.O.P.S.;		
RL	Submitted (SBP-2000) to the EMBL/GenBank/DBBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RA	Legoux, P., Legoux R., O'Brien D., Salome M.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBBJ databases.		
DR	EMBL: AJ299719; CACB2491; -.		
DR	GeneDB: 18374; HPS2.		
DR	InterPro: IPR005199; Glyco_hydro_79N.		
DR	Pfam: PF01662; Glyco_hydro_79N; T.		
SQ	SEQUENCE 592 AA; 66520 MW; 9498841FAC0558B CRC64;		

Query Match 40.3%; Score 1146.5; DB 4; Length 592; Best Local Similarity 43.4%; Pred. No. 6.3e-83; Matches 249; Conservative 82; Mismatches 190; Indels 53; Gaps 9;

RESULT 9

Q8WQ1	PRELIMINARY;	PRT;	548 AA.
AC	Q8WQ1;		
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DB	Heparase 3.		
OS	Homo sapiens (Human)		
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RC	Pessegue Safontas B.J.O.P.S.;		
RL	Submitted (SBP-2000) to the EMBL/GenBank/DBBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RA	Legoux, P., Legoux R., O'Brien D., Salome M.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBBJ databases.		
DR	EMBL: AJ299720; CACB2492; -.		
DR	InterPro: IPR005199; Glyco_hydro_79N.		
DR	Pfam: PF01662; Glyco_hydro_79N; T.		
SQ	SEQUENCE 548 AA; 61771 MW; B8986303FC73A60A CRC64;		

Query Match 35.7%; Score 1015.5; DB 4; Length 548; Best Local Similarity 41.8%; Pred. No. 1.7e-72; Matches 224; Conservative 79; Mismatches 180; Indels 53; Gaps 9;

RESULT 10

Q8WQ1	PRELIMINARY;	PRT;	548 AA.
AC	Q8WQ1;		
DT	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DB	Heparase 3.		
OS	Homo sapiens (Human)		
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RC	Pessegue Safontas B.J.O.P.S.;		
RL	Submitted (SBP-2000) to the EMBL/GenBank/DBBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Prostate;		
RA	Legoux, P., Legoux R., O'Brien D., Salome M.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBBJ databases.		
DR	EMBL: AJ299720; CACB2492; -.		
DR	InterPro: IPR005199; Glyco_hydro_79N.		
DR	Pfam: PF01662; Glyco_hydro_79N; T.		
SQ	SEQUENCE 548 AA; 61771 MW; B8986303FC73A60A CRC64;		

Query Match 35.7%; Score 1015.5; DB 4; Length 548; Best Local Similarity 41.8%; Pred. No. 1.7e-72; Matches 224; Conservative 79; Mismatches 180; Indels 53; Gaps 9;





Db	117	-----PFORM-----NSLGFSGSKGLHMRKWDDELSFL
Qy	187	GLNALLRTADLQ-----WNSNAQALLDXCSCSKGNY-SWELGNPNSF
Db	155	GLNALLRTADLQ-----WNSNAQALLDXCSCSKGNY-SWELGNPNSF
Qy	239	GSQLESDYIQLAKLRLSTFKNAKLGCPDVGSQP-----RKYAKMLKSLFLK
Db	213	KELYGKOLIVKDVINK-WKQSWLHKEILVARGFVQEQWMTKLEI-----
Qy	294	TWHHYKYLNGRT-----ATRDEDELPDVDFISSIONKVF-----QVNESTRGKK
Db	269	THHINNIGJGSGNDPALVKKIMPS-----YLSQVSKTFKDVNQTOHQHGWAS
Qy	348	YGGARPLISDTPAAGTMWFLDKLGLSARMGIEVMMRQFFGAGNYHVDEN
Db	325	YNSGRHRHVSDFPIDSFWYDOLGMSARHNTKTYCROTUV-GPYGLBLKGT
Qy	407	SLLFKLVLGKVIMMSAVGSGSKRKRLVYLHCTNTDNRYKEGDTLYAINL
Db	384	ALLWHLRMLGKQVLAQVTDGPP-----QLVRYAHCSK-----GRACVTLINL
Qy	465	-----RLPYPS-----NKGQVDTKLRR-----LGPHG
Db	436	VSNGINWVLMABESRKCKSLLDPTKURPFSWIGSKASDGYNLRNREBYHLTPENG
Qy	496	NGLTJLKVDDOTLPLMPEKPLRP-GSSLGLPAFSYSFPIRNKAAC
Db	496	NGKSILKPTATCDIPSU-BPVRSVNSPLNVLPLMSFIVLPNFDSAC
RESULT 15		
Q8H615		PRELIMINARY; PRT; 544 AA.
ID	Q8H615	
AC	Q8H615;	
DT	01-MAR-2003 (TREMBLrel. 23, Created)	
DT	01-MAR-2003 (TREMBLrel. 23, last sequence update)	
DR	OSJNBA0035103_11 protein.	
GN	OSJNBA0035103_11	
OS	Oryza sativa (Japonica cultivar-group).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryzeae; Oryzaeae; Oryza.	
OC	Embryophytae; Oryzeae; Oryza.	
OX	NCBI_TaxID=3947;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. Nipponbare;	
RA	Sasaki T., Matsumoto T., Yamamoto K.,	
RT	"Oryza sativa nipponbare (GA3)" genomic DNA, chromosome 6, BAC clone:OSJNBA0035103."	
RT	Submitted (DECA-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AP003019; BAC2238_1;	
DR	SEQUENCE 544 AA; 58723 MW; ECDB695F0B224269 CRC64;	
Query Match 13 8%; Score 392.5; DB 10; Length 54		
best Local Similarity 26.8%; Pred. No. 1e-22; Mismatches 179; Indels 15		
Matches 144; Conservative 58; Mismatches 1; Index 15		
Qy	75	LGSPIKURTLARGLSPYLRFGEKTKDPLIFD-----PKKESTPERS
Db	93	LSNKILNNAIRAFSPKLRUGLSIQLQKLYVGGGGCAPVNTNTSMFPG
Qy	126	ICKYSSIPDVKEKRLLEWPMQBOQILIREHYQKFKNSTYSSSDVYTF
Db	147	-----GOLPLH-----RW-----DELAFA
Qy	186	FGLNALLRTADL-----QWNSNAQALLDXCSCSKGNY-SWELGNPNS
Db	170	FGLNALLGRVPLPDSMGGPHDVTYASLIRYTASKGYKINGWLGNE-----
Qy	238	NGSOLGEDIYOLHKLLRKSTFKNAKLGCPDV-----GQPRKTA
Qy	:	:

LKKADIFIN 238  
 -SGVGASVS 212  
 AGEVIDSV 293  
 VVGETSSA 347  
 PWGEESGA 324  
 FDLRDIWL 406  
 |||||:  
 FVENRPDYS 383  
 HNTVKYL- 464  
 |||:  
 SNQSDFTVS 435  
 -LUSKSVOL 495  
 VLRSKTMVL 495  
 2  
 2  
 phyta;  
 4;  
 7; Gaps 19;  
 YWOSQVNQD 125  
 FTQ----- 146  
 ANCSGDLI 185  
 ||| :: 169  
 FLKADIFI 237  
 ------L 218  
 RQKSGARIV  
 RQKKSPLKA 285



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using swi model

Run on: October 22, 2003, 20:19:45

(without alignments)

1002.192 Million cell updates/sec

PI	Feinstein, B., Pecker, I., Vlodavsky, I.	ID	AB08849 standard; Protein; 543 AA.
XX		XX	
DR	WPI; 1999-302255/25.	AC	AB08849;
XX	N-PSDB; AAX35648.	XX	
PT	New human polynucleotide useful for treating angiogenesis, restenosis, and inflammation	DT	15-JAN-2001 (First entry)
XX		XX	
PS	Claim 6; Fig 1; 63pp; English.	DB	Amino acid sequence of a human heparanase polypeptide.
XX		XX	
CC	The specification describes a polypeptide having heparanase (hpa) activity. The recombinant protein is used as a modulator of heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoal and bacterial infections, or disintegration of neurodegenerative plaques. Heparanase may be useful for conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases, and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, and anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. The present sequence represents human heparanase.	KW	Human; heparanase; gene therapy; tumour; inflammation; autoimmunity; heparin-binding growth factor; cytokine; neurodegenerative plaque; wound healing; infection; burn; angiogenesis; restenosis; atherosclerosis; inflammation; neurodegenerative disease; Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
CC		XX	
CC		OS	Homo sapiens.
CC		XX	
CC		PN	WO200052178-A1.
CC		XX	
CC		PD	08-SEP-2000.
CC		XX	
CC		PR	14-FEB-2000; 2000WO-US03542.
CC		XX	
XX	Sequence 543 AA;	PR	01-MAR-1999; 99US-0258892.
XX		XX	
SQ	Query Match 100%; Score 2842; DB 20; Length 543; Best Local Similarity 100.0%; Pred. No. 9.5e-274; Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.
		PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.
		XX	
Db	1 MURSKPALPPMLLULGPGPSGALPRPAQAOVVDLDFPDKTDFLFDPKGESEPEERSWQS 60	DR	WPI; 2000-579289/54.
Db	1 MURSKPALPPMLLULGPGPSGALPRPAQAOVVDLDFPDKTDFLFDPKGESEPEERSWQS 60	XX	N-PSDB; AAX75051.
Db	1 IDANLATDPFLILGSPKURTLARGSPAYLRFGGTKTDFLFDPKGESEPEERSWQS 120	XX	
Db	1 IDANLATDPFLILGSPKURTLARGSPAYLRFGGTKTDFLFDPKGESEPEERSWQS 120	XX	
Qy	1 QNDICKYGSIPPDVKEKLEWYQDOLIREHYQKFKNSYSSSVVDLYTFANS 180	XX	
Qy	1 QNDICKYGSIPPDVKEKLEWYQDOLIREHYQKFKNSYSSSVVDLYTFANS 180	XX	
Db	181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSKSGKNTISWELGNBPNSPFLKKADIFINGS 240	XX	
Db	181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSKSGKNTISWELGNBPNSPFLKKADIFINGS 240	XX	
Qy	241 QLGHDYIQHLKLRSKTPKNAKLYGPDVQSPRRTAKOMLSFLKAGGEVIDSYWHHHL 300	XX	
Qy	241 QLGHDYIQHLKLRSKTPKNAKLYGPDVQSPRRTAKOMLSFLKAGGEVIDSYWHHHL 300	XX	
Db	301 NGGRATREDPLNPVUDIFSSVQKUFOVESTRPGKQWLGESSAYGGAPLSDTPA 360	XX	
Db	301 NGGRATREDPLNPVUDIFSSVQKUFOVESTRPGKQWLGESSAYGGAPLSDTPA 360	XX	
Qy	361 AGFMFLDKUGLSAMANGIEVMMRQPFAGGNYHLDENFDPLPDYWLSSLPKKGWGTKLM 420	XX	
Db	361 AGFMFLDKUGLSAMANGIEVMMRQPFAGGNYHLDENFDPLPDYWLSSLPKKGWGTKLM 420	XX	
Qy	421 ASVQSKRKRKVRVHLCTNTDNPVYKEGDLTLYINLHVKYRVLPPFSNKVQDKYLL 480	XX	
Db	421 ASVQSKRKRKVRVHLCTNTDNPVYKEGDLTLYINLHVKYRVLPPFSNKVQDKYLL 480	XX	
Qy	481 RPQGPHGLLSKSYQVQNGLTKMVDQDTPPLMELPRLPESQSSGAPAFSSEFVIRNAKA 540	XX	
Db	481 RPQGPHGLLSKSYQVQNGLTKMVDQDTPPLMELPRLPESQSSGAPAFSSEFVIRNAKA 540	XX	
Qy	541 ACI 543	Db	61 IDANLATDPFLILGSPKURTLARGSPAYLRFGGTKTDFLFDPKGESEPEERSWQS 120
Db	541 ACI 543	Qy	121 QNDICKYGSIPPDVKEKLEWYQDOLIREHYQKFKNSYSSSVVDLYTFANS 180
RESULT 2	AAB08849	Db	121 QNDICKYGSIPPDVKEKLEWYQDOLIREHYQKFKNSYSSSVVDLYTFANS 180



XX Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic; anti diabetic; immunomodulatory; anti-inflammatory; nephrotoxic; metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma; mesothelioma; melanoma; lymphoma; leukemia; cancer; sebosis; diabetes; inflammation; hemorrhagic nephritis; nephrotic syndrome, autoimmune disease; anticancer; kidney disease.  
 XX Homo sapiens.  
 XX WO957153-A1.  
 PD 11-NOV-1993.  
 XX 29-APR-1999; 99WO-US09255.  
 XX 01-MAY-1998; 98US-0071739.  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES. SERVICES & DEV.  
 PA (FRIE-) FRIEDMAN M M.  
 XX Pecker I, Vladavsky I, Friedman Y, Perets T;  
 PI DR WPI; 2000-052944/04.  
 DR N-PSDB; A4233290.  
 XX PT Heparanase-specific molecular probes useful for diagnosis and treatment, e.g. of tumors, and for targeted drug delivery  
 PS Example; Page 81-82; 90pp; English.  
 XX The present invention describes heparanase-specific molecular probes, useful for methods of detecting heparanase in situ. The probes and anti-heparanase antibodies are used to detect or quantify the expression of heparanase, for diagnosis and monitoring of diseases (especially metastasis), for treatment of heparanase-associated diseases (e.g. tumours, adenocarcinoma, squamous cell carcinoma, teratocarcinoma, mesothelioma, lymphoma or leukemia, a solid cancer or its metastases) derived from liver, prostate, bladder, breast, ovary, cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney disease, diabetes and inflammation syndrome, sepsis and inflammatory or autoimmune disease), for targeted drug delivery (e.g. of anticancer agents) and as research reagents. The present sequence represents human heparanase, which is used in the exemplification of the present invention.  
 XX Sequence 543 AA:

Query Match Best Local Similarity 100.0%; Score 2842; DB 21; Length 543;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKPALPPMLLGLPLGPGPLSPALPRAQADVVDLDFPTORPLHIVSPSLSVT 60  
 QY 1 MLLRSKPALPPMLLGLPLGPGPLSPALPRAQADVVDLDFPTORPLHIVSPSLSVT 60

QY 61 IDMLAATPPLTIGSKURTARGSPAYLRGGITDELFDPKKESTPERSYWQS 120  
 QY 61 IDMLAATPPLTIGSKURTARGSPAYLRGGITDELFDPKKESTPERSYWQS 120

QY 121 QNQDICKYGSIPDPVVEKRLWLPYQOBLLRBEHYKFKRKFSTYRSVVDYTTANCS 180  
 QY 121 QNQDICKYGSIPDPVVEKRLWLPYQOBLLRBEHYKFKRKFSTYRSVVDYTTANCS 180

QY 181 GHLIGLIGNALLRTADLQNSNSAQQLIDYCSKGYNTISWELGNEBNSFLKKIDITGS 240  
 QY 181 GHLIGLIGNALLRTADLQNSNSAQQLIDYCSKGYNTISWELGNEBNSFLKKIDITGS 240

QY 241 QLGEDYQLHKLKURKSTKNAKLYGPGQSPKTAQMLKSLKAGEEVISWTHYIL 300  
 QY 241 QLGEDYQLHKLKURKSTKNAKLYGPGQSPKTAQMLKSLKAGEEVISWTHYIL 300

QY 301 NGATATREDPLNPVLDIFISSVQKVEQVNESTRPCKQKWWLGTSAYGGAPLSDTFA 360  
 QY 301 NGATATREDPLNPVLDIFISSVQKVEQVNESTRPCKQKWWLGTSAYGGAPLSDTFA 360

Db 361 AGFWMLDKLQGSLSARGNIEVMYRQVFGAGNYHLDENFDPLPYWLSLFLPKLKGTKYL 420  
 Db 361 AGFWMLDKLQGSLSARGNIEVMYRQVFGAGNYHLDENFDPLPYWLSLFLPKLKGTKYL 420

QY 421 ASVQGSKRKRLRVYLHCTNTDNPYKEGDLTLAINLHVTKLRLPPFSNQKDVYL 480  
 QY 421 ASVQGSKRKRLRVYLHCTNTDNPYKEGDLTLAINLHVTKLRLPPFSNQKDVYL 480

Db 481 RPLGPHGLSKSVOLANGLIDKMDQTLPPMLMEKPLRPOSSLGIPAFSYSPFIRNAKVA 540  
 Db 481 RPLGPHGLSKSVOLANGLIDKMDQTLPPMLMEKPLRPOSSLGIPAFSYSPFIRNAKVA 540

QY 541 ACT 543  
 Db 541 ACT 543

RESULT 5  
 AAY97635  
 ID AAY97635 standard; Protein: 543 AA.  
 XX  
 AC AAY97635;  
 XX DT 20-APR-2001 (First entry)  
 DB Human: heparanase protein sequence.  
 XX  
 DB Human: heparanase protein sequence.  
 XX Heparanase; humpf; wound healing; angiogenesis; restenosis; Scrape; atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease; neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection; gene therapy; human.  
 XX Homo sapiens.  
 XX WO200100643-A2.  
 XX PD 04-JAN-2001.  
 XX PR 19-JUN-2000; 2000WO-IL00358.  
 XX PR 25-JUN-1999; 99US-0140801.  
 XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX PI Pecker I, Michal I, Itzhaki H;  
 DR WPI; 2001-137930/14.

XX PT New polynucleotides and polypeptides that are distantly homologous to heparanase, useful in wound healing, as well as in gene therapy protocols for angiogenesis, restenosis, atherosclerosis, or inflammation.

XX PS Disclosure; Page 64-65; 67pp; English.

XX This sequence represents a heparanase of the invention. The heparanase DNA and protein sequences are useful in wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary diseases, neurodegenerative diseases (such as Scrape, Alzheimer's disease, and Creutzfeldt-Jakob disease) or viral infections. The heparanase coding sequence is particularly useful in gene therapy.

XX Séquence 543 AA:

Query Match Best Local Similarity 100.0%; Score 2842; DB 22; Length 543;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1. MLLRSKPALPPMLLGLPLGPGPLSPALPRAQADVVDLDFPTORPLHIVSPSLSVT 60

DR	WPI; 2002-338926/37
PT	Nucleic acid encoding avian and reptile heparanase polypeptide 19
PT	useful to treat various heparin-related disorders and the signal
PT	peptide is useful in production of membrane-targeted or secreted
PT	recombinant proteins
XX	Disclosure; Fig 1a, 39pp; English.
PS	The invention relates to an isolated avian and reptile nucleic acid, encoding a polypeptide with heparanase catalytic activity. The signal peptide of the nucleic acid can be used to express membrane-associated or secreted proteins in heterologous expression systems. The encoded polypeptides can be used to prevent tumour angiogenesis, metastasis and invasion, and to intervene with pathologies associated with impaired heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoa and bacterial infections or disintegration of neurodegenerative plaques. The present sequence represents a human heparanase protein sequence used in similarity studies.
XX	Sequence 543 AA;
SQ	Query Match 100.0%; Score 2842; DB 23; Length 543; Best Local Similarity 100.0%; Pred No. 9.5e-24; Mismatches 0; Indels 0; Gaps 0; Matches 543; Conservative 0;
QY	1 MLLRKSPALPPMLLPGPLSPGALPRAQAOQDVDPDFQEPHLVSPSFLST
Db	1 MLLRKSPALPPMLLPGPLSPGALPRAQAOQDVDPDFQEPHLVSPSFLST 60
QY	1 DANIATDPRELLGSPKRTLARGSPAYLRFQGKTKDPLIFDKKESFEERYSWOS
Db	1 DANIATDPRELLGSPKRTLARGSPAYLRFQGKTKDPLIFDKKESFEERYSWOS 120
QY	121 OVNODICKYGSIPPVEKURLEWPMQBOULLREHFKKENSTSRSSVUVTYFANS
Db	121 OVNODICKYGSIPPVEKURLEWPMQBOULLREHFKKENSTSRSSVUVTYFANS 180
QY	181 GLDLIFGLNALLRTADLQWNSNAQOLLDYQSSKGYNISMIGNEPNSFLKKADIFINGS
Db	181 GLDLIFGLNALLRTADLQWNSNAQOLLDYQSSKGYNISMIGNEPNSFLKKADIFINGS 240
QY	241 OLGEDYIOLHKURKSTFKONKLYGPDVGORRKAKMUSPLKAGGEVIDSVTMHHYL
Db	241 OLGEDYIOLHKURKSTFKONKLYGPDVGORRKAKMUSPLKAGGEVIDSVTMHHYL 300
QY	301 NGRTATREDPFDLNDIFDIFISSVQVKFOVWSTRKPKWNGETSSAYGGAPLLSDTFA
Db	301 NGRTATREDPFDLNDIFDIFISSVQVKFOVWSTRKPKWNGETSSAYGGAPLLSDTFA 360
QY	361 AGFMWLDKLGISARMGIEVNRQVQFAGNTHLVDENFDPLPDWYSLFLKLUVTYFANS
Db	361 AGFMWLDKLGISARMGIEVNRQVQFAGNTHLVDENFDPLPDWYSLFLKLUVTYFANS 420
QY	421 ASVOGSKERKURVYIACNTNDPRTKEGDLTLYAINHNVTKYLRYPFPNSKQVDKLL
Db	421 ASVOGSKERKURVYIACNTNDPRTKEGDLTLYAINHNVTKYLRYPFPNSKQVDKLL 480
QY	481 RPLPGLGLSLSSVQNLGLTQVMDQDPLPPLMEKPLRPSSGLGPAPSYSFVIRNAKVA
Db	481 RPLPGLGLSLSSVQNLGLTQVMDQDPLPPLMEKPLRPSSGLGPAPSYSFVIRNAKVA 540
QY	541 ACI 543
Db	541 ACI 543
RESULT 6	
ABB07813	ABB07813 standard; protein; 543 AA.
XX	
AC	ABB07813;
XX	
DT	03-JUL-2002 (first entry)
DB	
XX	Human heparanase sequence.
KW	Reparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme; anti-protozoan; neuroprotective; heparin; human.
OS	Homo sapiens.
XX	
PH	Key Location/Qualifier
PT	1..39 /note= "signal peptide"
PT	36..543 /note= "mature protein"
XX	
PN	US2002034810-A1.
XX	21-MAR-2002.
PR	16-AUG-2001; 2001US-0930218.
XX	
XX	20-SEP-2000; 2000US-0666390.
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX	
PI	Goldshmidt O, Pecker I, Vladavsky I, Michal I, Zcharia B;
XX	
RESULT 7	
ARY02346	ID AAY02346 standard; Protein; 592 AA.
XX	

AC AAY02346;  
 XX DT 09-JUL-1999 (first entry)  
 XX DR A human heparanase protein.  
 XX KW Heparanase; hpa; modulator; heparin-binding growth factor;  
 KW cellular response; cytokine; cell interaction; plasma lipoprotein;  
 KW cellular susceptibility; infection; disintegration;  
 KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;  
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.  
 OS Homo-sapiens.  
 XX DN WO991798-A1.  
 XX ACK 11-MAR-1999.  
 XX PR 31-AUG-1998; 98WO-US17954.  
 XX PR 02-JUL-1998; 98US-0109386.  
 XX PR 02-SEP-1997; 97US-0922170.  
 PA (FRIE/ ) FRIEDMAN M. M.  
 PA (HADA/ ) HADASIT MEDICAL RES SERVICES & DEV.  
 XX (INSI/ ) INSIGHT STRATEGY & MARKETING LTD.  
 PR Feinstein E, Pecker I, Vlodavsky I;  
 XX DR N-PSDB; AAX35650.  
 XX PT New human polynucleotide useful for treating angiogenesis,  
 PT restenosis, and inflammation  
 PS Claim 6; page 65-66; 63pp; English.  
 XX The specification describes a polypeptide having heparanase (hpa) activity. The recombinant protein is used as a modulator of heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoal and bacterial infections, or disintegration of neurodegenerative plaques. Heparanase may be useful for conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases, and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, and anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. The present sequence represents human heparanase.  
 XX Sequence 592 AA:  
 Query Match 100.0%; Score 2842; DB 20; Length 592;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-273; Mismatches 0; Indels 0; Gaps 0;  
 Matches 543; Conservative 0; MisMatch 0; Indel 0;  
 QV 1 MULRSKPAUPPMUMLLPGPLGSPGALPRPAQDQVVDLDPFTOPBLHVLVSPRSLSVT 60  
 DB 50 MULRSKPAUPPMUMLLPGPLGSPGALPRPAQDQVVDLDPFTOPBLHVLVSPRSLSVT 109  
 61 IMLNLTDPFLTIGSPKRTARGSPAVYRFGSKTDELFDPKESTEERSTWS 120  
 110 IMLNLTDPFLTIGSPKRTARGSPAVYRFGSKTDELFDPKESTEERSTWS 169  
 121 QWQDICKYGSPPDVEBKRLRMEWPKOBLLRERHKKKRSTYSSSVVLYTFRANS 180  
 DB 170 QWQDICKYGSPPDVEBKRLRMEWPKOBLLRERHKKKRSTYSSSVVLYTFRANS 229  
 QY 181 GUDLIGFQNALRLLTADLQNSNSAQLLIDYCSSSKGYNTSWEGLNEPNSFLKQDIFNGS 240  
 DB 230 GUDLIGFQNALRLLTADLQNSNSAQLLIDYCSSSKGYNTSWEGLNEPNSFLKQDIFNGS 289

QY 241 QLGEDYIOLKLKLKSTPKNAKLGDPVGOPRRTAKOMLSFLKAGSEVIDSVTWHLYL 300  
 DB 290 QLGEDYIOLKLKLKSTPKNAKLGDPVGOPRRTAKOMLSFLKAGSEVIDSVTWHLYL 349  
 DR 301 NGRTATREDFELNPVLDIFISSVQKUQFQESTRPGKOMLGETSSAVGGAPLSDPFA 360  
 DR 350 NGRTATREDFELNPVLDIFISSVQKUQFQESTRPGKOMLGETSSAVGGAPLSDPFA 409  
 QY 361 AGFWMLDKLGSARMGIEVMRQFFGAGNYLAVDENDPDPOWLSLFLKLGVLGKTLM 420  
 DR 410 AGFWMLDKLGSARMGIEVMRQFFGAGNYLAVDENDPDPOWLSLFLKLGVLGKTLM 469  
 QY 421 ASVOCISKRKURVLTCHCTMDNPRYKECDLTLA1LNHNTKYLRLPPPSNQVDKYL 480  
 DR 470 ASVOCISKRKURVLTCHCTMDNPRYKECDLTLA1LNHNTKYLRLPPPSNQVDKYL 529  
 DR 481 RPLGHGLIISKSQVQNLGJLKMVDQDTPPLMPLRGCSSLGPAVSYSEVIRNAKA 540  
 DR 530 RPLGHGLIISKSQVQNLGJLKMVDQDTPPLMPLRGCSSLGPAVSYSEVIRNAKA 589  
 QY 541 ACI 543.  
 DR 590 ACI 592

RESULT 8  
 AAB08850 standard; Protein: 592 AA.  
 ID AAB08850  
 XX AC AAB08850;  
 XX DT 15-JAN-2001 (first entry)  
 XX DE Amino acid sequence of a human heparanase polypeptide.  
 XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity; heparin-binding growth factor; cytokine; neurodegenerative plaque; wound healing; infection; burn; angiogenesis; restenosis; atherosclerosis; inflammation; neurodegenerative disease; Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
 XX OS Homo sapiens.  
 XX PN WO200052178-A1.  
 XX PD 08-SEP-2000.  
 XX PP 14-FEB-2000; 2000WO-US03542.  
 XX PR 01-MAR-1999; 99US-0258892.  
 XX PA (INSI/ ) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA/ ) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/ ) FRIEDMAN M. M.  
 XX PI Pecker I, Vlodavsky I, Feinstein E;  
 XX DR WPI; 2000-579289/54;  
 DR N-PSDB; AAX75053.  
 XX New polynucleotides encoding a polypeptide having heparanase activity, useful in wound healing and in gene therapy, particularly in treating tumour, inflammation, autoimmunity, neurodegenerative diseases particularly, the polynucleotide is useful in modulating the bioavailability of heparin-binding growth factors, cellular responses to heparin-binding growth factors (e.g. bFGF) and cytokines



QY .481 RPLGPHGLSKSVQNLGLTRKMDOTLPPMEKPLRGSSIGLPAFSYSPFVIRNAKA 540  
 QY .481 RPLGPHGLSKSVQNLGLTRKMDOTLPPMEKPLRGSSIGLPAFSYSPFVIRNAKA 540  
 QY .541 ACT 543  
 Db 541 ACT 543

RESULT 10  
 ID AAB86206 standard; Protein: 543 AA.  
 XX  
 AAB86206;  
 XX  
 DT 24-AUG-2001 (first entry)  
 DB Human heparanase inhibitor protein.  
 XX  
 KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic; hepatotropic; veterinary medicine; congestive heart failure; dyspnoea; primary cardiomyopathy; peripheral edema; pulmonary congestion; hepatic congestion; hydrothorax; ascite; nocturia; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19955803-A1.  
 XX  
 PD 23-MAY-2001.  
 XX  
 PP 19-NOV-1999; 99DE-1055803.  
 PR 19-NOV-1999; 99DE-1055803.  
 XX  
 PA (KNOL ) KNOLL AG.  
 XX  
 PI Herr D, Hahn A, Laux V;  
 XX  
 DR WPI; 2001-368371/39.  
 DR N-PSDB; AAH20940.  
 PT Treatment or prevention of cardiac insufficiency and related conditions, e.g. pulmonary congestion and dyspnoea, comprises administration of heparanase inhibitor  
 XX  
 PS Disclosure; Page 11-13; 16pp; German.  
 CC This invention describes a novel heparanase inhibitor which can be used for the treatment or prevention of cardiac insufficiency and associated indications, symptoms and/or malfunctions. The heparanase inhibitor of the invention has cardiant, nephrotropic and hepatotropic activity. The products of the invention can be used in human and veterinary medicine, for the treatment or prevention of congestive heart failure e.g. primary cardiomyopathy. Associated conditions treated or prevented with the inhibitor are especially peripheral edemas, pulmonary and hepatic congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g. nocturia can also be treated. This sequence represents the human heparanase protein described in the method of the invention.  
 XX  
 SQ Sequence 543 AA;

Query Match 99.9%; Score 2838; DB:22; Length 543;  
 Best Local Similarity 99.8%; Pred: No 2, 4e-273; Mismatches 1; Indels 0; Gaps 0;

QY .121 QVNQICKCGGSIPPVEVKRLRMPYQQLLRLRHYQKPKKNTSYRSVSDVLYTFANCS 180  
 QY .121 QVNQICKCGGSIPPVEVKRLRMPYQQLLRLRHYQKPKKNTSYRSVSDVLYTFANCS 180  
 QY .181 GLDIFGLNLRLRPLQNSNQQLDYCQSSKGNTSWELNEPNSFLKKDIFNGS 240  
 Db 181 GLDIFGLNLRLRPLQNSNQQLDYCQSSKGNTSWELNEPNSFLKKDIFNGS 240  
 QY .241 OLGEDYQIQLKLLRKGSTPKQAKLYGPDVQCPQPKTAKLKSFLKAGGIVDSTWHRYL 300  
 Db 241 OLGEDYQIQLKLLRKGSTPKQAKLYGPDVQCPQPKTAKLKSFLKAGGIVDSTWHRYL 300  
 QY .361 AGFMWLDKLGSLSANGIENYVWROFFGAGNYHLYDENFDPLPYWLSLFLKKLUGTKLM 420  
 Db 361 AGFMWLDKLGSLSANGIENYVWROFFGAGNYHLYDENFDPLPYWLSLFLKKLUGTKLM 420  
 QY .421 ASVQSKRKRKRVVLAHTMDTNPYKEDGDTIKAINTAKTLYLKPYPPEFSNKQDVYL 480  
 Db 421 ASVQSKRKRKRVVLAHTMDTNPYKEDGDTIKAINTAKTLYLKPYPPEFSNKQDVYL 480  
 QY .481 RPLGPHGLSKSVQNLGLTRKMDOTLPPMEKPLRGSSIGLPAFSYSPFVIRNAKA 540  
 Db 481 RPLGPHGLSKSVQNLGLTRKMDOTLPPMEKPLRGSSIGLPAFSYSPFVIRNAKA 540  
 QY .541 ACT 543  
 Db 541 ACT 543

RESULT 11  
 ID AAY30124  
 ID AAY30124 standard; Protein: 588 AA.  
 XX  
 AAY30124;  
 XX  
 DT 20-MAR-2003 (updated)  
 DT 14-OCT-1999 (first entry)  
 XX  
 DE A human protein with heparanase activity.  
 XX  
 KW Human; heparanase; heparan sulfate; trauma; autoimmune disease; skin disease; cardiovascular disease; nervous system disease; Alzheimer's disease; cancer; cancer metastasis; angiogenesis; inflammation; arthritis.  
 OS Homo sapiens.  
 XX  
 PN WO9940207-A1.  
 XX  
 PR 12-AUG-1999;  
 PR 05-FEB-1999; 99WO-EP00777.  
 XX  
 PR 09-FEB-1998; 98GB-0002725.  
 XX  
 PA (NOV ) NOVARTIS AG.  
 PA (NOV ) NOVARTIS-ERINDUNGEN VERW GES MBH.  
 XX  
 PA Nakajima M, Toyosima M;  
 XX  
 DR WPI; 1999-494300/41.  
 N-PSDB; AAX86671.  
 XX  
 PT New heparanase polypeptide useful for treating autoimmune diseases, skin diseases, cardiovascular diseases and nervous system diseases including Alzheimer's disease

QY .61 IDNLANDPPLLIGSPKLRGSPAYKRGKTKDFLIDPKKESTPERSYQWS 120  
 QY .61 IDNLANDPPLLIGSPKLRGSPAYKRGKTKDFLIDPKKESTPERSYQWS 120  
 Ps Claim 3; Page 29-31; 40pp; English.

xx

The present sequence represents a polypeptide with human heparanase biological activity. Antagonists and inhibitors of the protein prevent it from degrading the extracellular matrix and releasing heparan sulfate

from the extracellular matrix surface. The heparanase protein or the anti-heparanase antibody are used in pharmaceutical compositions for treating warm blooded animals suffering from a disease resulting from shortage or lack of the heparanase protein, or from excessive activity or over-expression of the heparanase protein, respectively. The heparanase protein is used in treating diseases such as trauma, autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

xx

autoimmune disease, skin disease, cardiovascular diseases and nervous system diseases including Alzheimer's disease resulting from shortage or lack of polypeptide. The anti-heparanase antibody is used in treating the diseases like cancer, cancer metastasis, angiogenesis and inflammation including arthritis resulting from excessive activity or over-expression of heparanase protein. The anti-heparanase antibody can be used to detect the presence or absence of polypeptide and its concentration. (Updated on 20-MAR-2003 to correct PA field.)

23-MAY-2001 (first entry)  
DT  
xx  
DB Human membrane or secretory protein clone PSEC0090.  
xx  
KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
rheumatoid arthritis; diabetes.  
xx  
OS Homo sapiens.  
xx  
EP1067182-A2.  
xx  
EP1067182-A2.  
xx  
PD 10-JAN-2001.  
xx  
PR 07-JUL-2000; 2000BP-0114090.  
xx  
PR 08-JUL-1999; 99JP-0194179.  
xx  
PR 11-JAN-2001; 2000JP-0118775.  
xx  
PR 02-MAY-2000; 2000JP-0183766.  
xx  
PA (HLI-) HELIX RES INST.  
xx  
PI Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T., Hayashi, K.  
xx  
DR WPI; 2001-033989/11.  
xx  
N-PSDB; AAF33788.  
xx  
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development.  
xx  
PS Claim 1; SEQ ID 90; 609pp + CD ROM; English.  
xx  
This invention relates to nucleic acid sequences AAF93744 - AAF93916 which encode human secretory or membrane proteins represented by AAF88317 - AAF88119. Included in the invention are primers AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences and the production of antibodies directed against the proteins, and the diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.

Sequence 588 AA:  
Query Match Best Local Similarity 99.4%; Score 2838; DB 20; Length 588; Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
1 MLRSKPALPPMLLQPGPSPGALPRAQDQVUDLRFPTQBLHVSPLSFTV 60  
46 MLRSKPALPPMLLQPGPSPGALPRAQDQVUDLRFPTQBLHVSPLSFTV 105  
61 IDANLATAFPLILGSPKRLTARGSPAYLRFGGTKTDLFDPKKESTBERSWQS 120  
106 MLRSKPALPPMLLQPGPSPGALPRAQDQVUDLRFPTQBLHVSPLSFTV 165  
121 QVNDICKYGSIPPVEEKRLRPLWYQEBOLRHYQKPKNSTYRSSSDVLYTFCNS 180  
166 QVNDICKYGSIPPVEEKRLRPLWYQEBOLRHYQKPKNSTYRSSSDVLYTFCNS 225  
226 GLDIFGFLNALLRQLDQMLNSNQQLDLYCSSGYNISHELGEPNSLKKDIFING 265  
241 QLGDRYIQLKQLRKESTPKRAKLGDPVQSPRRTAKMRSFLAGGEVIDSVMWHYL 300  
286 QLGDRYIQLKQLRKESTPKRAKLGDPVQSPRRTAKMRSFLAGGEVIDSVMWHYL 345  
301 NGRTATREDTNPVDPIDISSVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 360  
346 NGRTATREDTNPVDPIDISSVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 405  
361 AGFWMLDKGGLSARNGIEVWVNRQPGAGSYHLDENFDPDLYWLSLFLKKGTKV 420  
406 AGFWMLDKGGLSARNGIEVWVNRQPGAGSYHLDENFDPDLYWLSLFLKKGTKV 465  
421 ASVQSKRKLRLVYHCTNTDNPVDPIDISSVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 480  
466 ASVQSKRKLRLVYHCTNTDNPVDPIDISSVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 525  
481 RPLGPHGLISKVQVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 540  
526 RPLGPHGLISKVQVQVQESTPKCKWLGESSAYGGAPLSDTFA 585  
541 ACI 543  
586 ACI 588  
Sequence 543 AA:  
Query Match Best Local Similarity 99.4%; Score 2826; DB 22; Length 543; Matches 540; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 MLRSKPALPPMLLQPGPSPGALPRAQDQVUDLRFPTQBLHVSPLSFTV 60  
61 IDANLATAFPLILGSPKRLTARGSPAYLRFGGTKTDLFDPKKESTBERSWQS 120  
61 IDANLATAFPLILGSPKRLTARGSPAYLRFGGTKTDLFDPKKESTBERSWQS 160  
121 QVNDICKYGSIPPVEEKRLRPLWYQEBOLRHYQKPKNSTYRSSSDVLYTFCNS 180  
121 QVNDICKYGSIPPVEEKRLRPLWYQEBOLRHYQKPKNSTYRSSSDVLYTFCNS 180

QY 181 GLDLIGLGINALLRTADIQWNSNAQQLDYCSSKGYNISWLGNEPNSFLKKADIFINGS 240  
 CC |||||  
 CC 181 GLDLIGLGINALLRTADIQWNSNAQQLDYCSSKGYNISWLGNEPNSFLKKADIFINGS 240  
 CC |||||  
 CC 241 QI~~G~~DYVQLHKLURKSTPKNAKLYGPDVGQRRKTAQMSPLKAGGEVIDSVTWHY 300  
 CC |||||  
 CC 241 QL~~G~~EPFIOLHKLURKSTPKNAKLYGPDVGQRRKTAQMSPLKAGGEVIDSVTWHY 300  
 CC |||||  
 CC 301 NGRTATREDFLNPVDIPLFISVQKFQVWVSTREPKKVMGETSSAYGGAPLISDTFA 360  
 CC |||||  
 CC 301 NGRTATREDFLNPVDIPLFISVQKFQVWVSTREPKKVMGETSSAYGGAPLISDTFA 360  
 CC |||||  
 QY 361 AGFMWFLDKLGLSARMGIEVWVQVFGAGNTHLVDENFDPLDYLWISLFLKKLGVTYU 420  
 CC |||||  
 CC 361 AGFMWFLDKLGLSARMGIEVWVQVFGAGNTHLVDENFDPLDYLWISLFLKKLGVTYU 420  
 CC |||||  
 QY 421 ASVQGSKERKURKVYLICNTNTPRVEGDTIYLAINHNVYKYLRLYPFPSENQKVDYU 480  
 CC |||||  
 CC 421 ASVQGSKERKURKVYLICNTNTPRVEGDTIYLAINHNVYKYLRLYPFPSENQKVDYU 480  
 CC |||||  
 QY 481 RPLPLGPGILSKSVQMLGTLKWDQDQTLPPIMEKPIRPQGSSGLPFLAFSSFPVIRNAKA 540  
 CC |||||  
 CC 481 RPLPLGPGILSKSVQMLGTLKWDQDQTLPPIMEKPIRPQGSSGLPFLAFSSFPVIRNAKA 540  
 QY 541 ACI 543  
 DB 541 ACI 543

RESULT 13

ABP56822  
 ID ABP56822 standard; Protein: 545 AA.  
 XX  
 AC ABP56822;  
 XX  
 DT 02-APR-2003 (first entry)  
 DB Human heparanase protein SEQ ID NO:18.  
 XX  
 KW cytotoxic; gene therapy; tumour; enzyme.  
 XX  
 OS Homo sapiens.  
 PN WO200304705-A1.  
 XX  
 PD 16-JAN-2003.  
 PP 01-JUL-2002; 2002WO-US20636.  
 XX  
 PR 05-JUL-2001; 2001US-089940.  
 XX  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Stein C;  
 XX  
 DR WPI; 2003-201558/19.  
 XX  
 N-PSDB; AB222816;

PT New oligonucleotide having a sequence complementary to a sequence of ribonucleic acid encoding a heparanase, useful for preparing a composition for treating tumor

PT Disclosure; Page 46-47; 48pp; English.

XX  
 CC The present invention describes an oligonucleotide having a sequence complementary to a sequence of ribonucleic acid encoding a heparanase. The oligonucleotide hybridises with the ribonucleic acid under conditions of high stringency and has a sequence comprising 10-10 bp. The oligonucleotide comprises at least one internucleotide linkages or the oligonucleotide comprise a phosphorothioate linkage. Hybridisation of the oligonucleotide to the ribonucleic acid inhibits expression of the heparanase, where inhibition

CC of heparanase means at least a 50% reduction in the quality of heparanase. Also described: (1) a method of inhibiting expression of a heparanase in a cell; (2) a composition comprising the above oligonucleotide in an amount effective to inhibit the expression of heparanase in the cell; and (3) a method of treating a tumour in a subject comprising administering to the subject an amount of the above oligonucleotide effective to inhibit expression of a heparanase in the subject. Heparanase antisense oligonucleotides have cytostatic activity, can be used in gene therapy, and can be used for preparing a composition for treating tumours. The present sequence represents human heparanase, which is given in the exemplification of the present invention.

CC  
 CC Sequence 545 AA:  
 SQ Query Match Similarity 99.1%; Score 2817; DB 24; Length 545;  
 Best Local Similarity 99.4%; Prod. No. 3e-271; Mismatches 0; Indels 2; Gaps 2;  
 Matched 542; Conservative 1; MisMatched 0;  
 QY 1 MLRSKPALEPP-IMLULIGPLCPSPALPRAQA-OVDVVIDDFTQBLHLVSPSPLS 58  
 DB 1 MLRSKPALEPP-IMLULIGPLCPSPALPRAQA-OVDVVIDDFTQBLHLVSPSPLS 60  
 QY 59 VTIQANLADTPRELIIGSPKURTLARGISPAVLRFGSTKTDLIEPKKESTPEERSW 118  
 DB 61 VTIQANLADTPRELIIGSPKURTLARGISPAVLRFGSTKTDLIEPKKESTPEERSW 120  
 QY 119 QSQVNQDICKYGSIPPDEVEKLAEBWPQOBQLLAREHYQKFKNSTRSVQLYTFAN 178  
 DB 121 QSQVNQDICKYGSIPPDEVEKLAEBWPQOBQLLAREHYQKFKNSTRSVQLYTFAN 180  
 QY 179 CSGDLIGLGINALLRTADIQWNSNAQQLDYCSSKGYNISWLGNEPNSFLKKADIFIN 238  
 DB 181 CSGDLIGLGINALLRTADIQWNSNAQQLDYCSSKGYNISWLGNEPNSFLKKADIFIN 240  
 QY 239 GSQLGEDYVQLHKLURKSTPKNAKLYGPDVGQRRKTAQMSPLKAGGEVIDSVTWHY 298  
 DB 241 GSQGEDYVQLHKLURKSTPKNAKLYGPDVGQRRKTAQMSPLKAGGEVIDSVTWHY 300  
 QY 299 YNGRTATREDFLNPVDIPLFISVQKFQVWVSTREPKKVMGETSSAYGGAPLISDT 358  
 DB 301 YNGRTATREDFLNPVDIPLFISVQKFQVWVSTREPKKVMGETSSAYGGAPLISDT 360  
 QY 359 FAAGFMWFLDKLGLSARMGIEVWVQVFGAGNTHLVDENFDPLDYLWISLFLKKLGVTYU 418  
 DB 361 FAAGFMWFLDKLGLSARMGIEVWVQVFGAGNTHLVDENFDPLDYLWISLFLKKLGVTYU 420  
 QY 419 LMASVQGSKERKURKVYLICNTNTPRVEGDTIYLAINHNVYKYLRLYPFPSENQKVDYU 478  
 DB 421 LMASVQGSKERKURKVYLICNTNTPRVEGDTIYLAINHNVYKYLRLYPFPSENQKVDYU 480  
 QY 479 LRLPLGPGILSKSVQMLGTLKWDQDQTLPPIMEKPIRPQGSSGLPFLAFSSFPVIRNAKA 538  
 DB 481 LRLPLGPGILSKSVQMLGTLKWDQDQTLPPIMEKPIRPQGSSGLPFLAFSSFPVIRNAKA 540  
 QY 539 VAACI 543  
 DB 541 VAACI 545

RESULT 14

AYA34173  
 ID AAYA34173 standard; Protein: 530 AA.  
 XX  
 AC AAYA34173;  
 XX  
 DT 15-NOV-1999 (first entry)  
 DB Human pre-heparanase protein sequence.

XX  
 CC Human; pre-heparanase; platelet; wound healing; angiogenesis blocker; inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis; heparin degradation; anticoagulant neutralisation; asthma; CNS disease; KW

**KW** inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;  
**KW** tumour growth; fibroproliferative disorder; neurodegenerative disease;  
**KW** therapy;  
**XX**  
**OS** Homo sapiens.  
**PN** WO9943830-A2.  
**XX**  
**PD** 02-SEP-1999.  
**XX**  
**PR** 18-FEB-1999; 99WO-US01489.  
**XX**  
**PR** 26-MAR-1998; 98US-0079401.  
**PR** 24-FEB-1998; 98US-0075706.  
**XX**  
**PA** (PHARM ) PHARMACIA & UPJOHN CO.  
**XX**  
**PI** Fairbanks MB, Heinrikson RL, Milderer AM;  
**XX**  
**DR** WPI; 1999-540598/45.  
**DR** N-PSDB; AAZ11236.  
**XX**  
**PT** New isolated platelet heparanase polypeptides, used to develop  
**PT** products for, e.g. wound healing and blocking angiogenesis  
**XX**  
**PS** Claim 12; Fig 7; 57pp; English.  
**XX**  
**CC** This sequence is the human pre-proheparanase of the invention. This  
**CC** sequence was isolated from human platelets. The heparanase can be used  
**CC** for identifying agents which alter heparanase activity. The heparanase  
**CC** can be used for wound healing or for blocking angiogenesis or  
**CC** inflammation. It can be used for treating e.g. psoriasis, diabetic  
**CC** retinopathy or solid tumours, or for the degradation of heparin and the  
**CC** neutralization of heparin's anticoagulant properties during surgery.  
**CC** Inhibitors of heparanase activity can be used in the treatment of  
**CC** atherosclerosis, asthma, and other inflammatory disease, vascular restenosis,  
**CC** disorders, and central nervous system (CNS) and neurodegenerative  
**CC** diseases. The products can also be used for detection and diagnosis. The  
**CC** purified heparanase, both recombinantly produced human heparanase and  
**CC** heparanase isolated from human platelet activity, allows for the  
**CC** convenient selection of compounds having anti-heparanase activity,  
**CC** i.e. inhibitors of heparanase activity, by measuring inhibition of  
**CC** heparanase activity. Inhibition of heparanase activity can be measured by  
**CC** blocking heparanase-mediated release of radioactive fragments from in  
**CC** vivo radiolabelled (H<sup>3</sup>PG) heparin.  
**XX**  
**SQ** Sequence 530 AA;

**Query** Match 97.3%; Score 2764; DB 20; Length 530;  
**Best** Local Similarity 99.4%; Pred. No. 5.4e-266;  
**Matches** 527; Conservat. 2; Mismatches 1; Indels 0; Gaps 0;

**QY** 14 MLLLGLPLGPSPGALPRAQDQDVLDPEPQEPHLVSPRSVITDANLADPFLI 73  
**Db** 1 MLLLGLPLGPSPGALPRAQDQDVLDPEPQEPHLVSPRSVITDANLADPFLI 60  
**QY** 74 LIGSPGRTWGLSPLASPLRGGTKYDPLFDPKKGESTERRSRWSQVNDICKYGSIP 133  
**Db** 61 LIGSPGRTWGLSPLASPLRGGTKYDPLFDPKKGESTERRSRWSQVNDICKYGSIP 120  
**QY** 134 PDPVERGLRPLPYQDQLLRHEHYQKFKFNSYTSRSVVDVLYTANCSGGLDIFUNALR 193  
**Db** 121 PDPVERGLRPLPYQDQLLRHEHYQKFKFNSYTSRSVVDVLYTANCSGGLDIFUNALR 180  
**QY** 194 TADLQMNNSQVLLDLYCQSSKGKNSWELGHBPNSLKADIPNSQGHDYQKHL 253  
**Db** 181 TADLQMNNSQVLLDLYCQSSKGKNSWELGHBPNSLKADIPNSQGHDYQKHL 240  
**QY** 254 RRSSTPKNAKLYGPPDVSOPRRTKAMKLSPLKAGEGEVDSVTHHYTNGTATREDPLNP 313  
**Db** 241 RRSSTPKNAKLYGPPDVSOPRRTKAMKLSPLKAGEGEVDSVTHHYTNGTATREDPLNP 300

**QY** 314 DVLDIFTSVVKQVQFWESTPREGKQWKGETTSAYGGAPLSDTAAGEMWMDLGSA  
**Db** 301 DVLDIFTSVVKQVQFWESTPREGKQWKGETTSAYGGAPLSDTAAGEMWMDLGSA 360  
**QY** 374 RMLGLEYVWVQPGAGMHLVUDENFDLPLDPMWLSLFLKKVJGKUWMSAGSGSKRKLV 433  
**Db** 361 RMGIFEVVNRQVFFGAGNTHLVDENFDLPLDPMWLSLFLKKVJGKUWMSAGSGSKRKLV 420  
**QY** 434 YLHCTNTDNPYKEGDLTLYAHLHNTKYLRLPYSNSKQVQYKYLRLPQPHGLSKV 493  
**Db** 421 YLHCTNTDNPYKEGDLTLYAHLHNTKYLRLPYSNSKQVQYKYLRLPQPHGLSKV 480  
**QY** 494 QLNGITLKVNDOTLPRIMEKPLRGPGSSLGAPSPSFRVTRNKAACI 543  
**Db** 481 QLNGITLKVNDOTLPLPKLPGSSLGAPSPSFRVTRNKAACI 530

**RESULT** 15  
**AA** YAY17083  
**ID** AAY17083 standard; Protein; 532 AA.  
**XX**  
**AC** AAY17083;  
**XX**  
**DT** 21-JUL-1999 (first entry)  
**XX**  
**DE** Seq ID No: 15 of WO9921975.  
**XX**  
**KW** Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
**KW** metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
**KW** arteriosclerosis; atherosclerosis; inflammation; tissue development;  
**KW** human; HSPG.  
**OS** Homo sapiens.  
**XX**  
**PN** WO9921975-A1.  
**XX**  
**PD** 06-MAY-1999.  
**XX**  
**PF** 28-OCT-1998; 99WO-AU00898.  
**XX**  
**PR** 09-DEC-1997; 97AU-0000812.  
**PR** 28-OCT-1997; 97AU-0000062.  
**XX**  
**PA** (AUUS ) UNIV AUSTRALIAN NAT.  
**XX**  
**PT** Freeman CG, Hamdorf BJ, Hullett MD, Parish CR;  
**XX**  
**DR** WPI; 1999-312956/26.  
**XX**  
**DR** N-PSDB; AAX37260.  
**XX**  
**PT** Polynucleotides encoding mammalian endoglucuronidases, especially  
**PT** heparanases, useful to promote wound healing  
**XX**  
**PS** Claim 6; Page 76-79; 112pp; English.

**XX**  
**CC** The invention relates to nucleic acid sequences that encode heparanase  
**CC** enzymes having endoglucuronidase activity. Recombinant heparanases are  
**CC** capable of removing the HS side chain from heparan sulfate proteoglycan  
**CC** (HSPG). Sulfated oligosaccharides, sulfonates or HSPG can be used to  
**CC** inhibit heparanase, this is useful for treatment of a physiological or  
**CC** medical condition associated with elevated heparanase activity, such as  
**CC** metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
**CC** arteriosclerosis, atherosclerosis and inflammation. The human murine and  
**CC** rat heparanases can be used to enhance wound healing, especially  
**CC** associated with tissue development and repair. The conditions mentioned  
**CC** above can be diagnosed using specific antibodies, and also using primers  
**CC** and probes specific for the heparanase polynucleotides. Other uses of the  
**CC** heparanases include sequencing sulfated molecules such as HSPG.  
**XX**  
**SQ** Sequence 532 AA;

**Query** Match 96.3%; Score 2737; DB 20; Length 532;  
**Best** Local Similarity 99.8%; Pred. No. 2.7e-263;

Matches	522;	Conservative.	1;	Mismatches	0;	Indels	0;	Gaps
Qy	1	MILRSKPALPPMLLGPGLPSRGALPRPAQADWVLDFFTOBPLHVLSSPLST	60					
Db	1	MILRSKPALPPMLLGPGLPSRGALPRPAQADWVLDFFTOBPLHVLSSPLST	60					
Qy	61	IDANLAATDPRFLILGSPKURTARGLSPAYLRFGCTKTOPKPKSTFERSYWS	120					
Db	61	IDANLAATDPRFLILGSPKURTARGLSPAYLRFGCTKTOPKPKSTFERSYWS	120					
Qy	121	QVNQDICKYSSIPDVEEKURLEWPTQEQLRERYQKKEKNSTYRSSDVWITFAMCS	180					
Db	121	QVNQDICKYSSIPDVEEKURLEWPTQEQLRERYQKKEKNSTYRSSDVWITFAMCS	180					
Qy	181	GLDLIFGLNALLRTADLQWNSNAQILLYCSSKGYNISBLGIPNSFLKKADIFINGS	240					
Db	181	GLDLIFGLNALLRTADLQWNSNAQILLYCSSKGYNISBLGIPNSFLKKADIFINGS	240					
Qy	241	QLGEDYIOLHKLRLKSTFKNAKLYGSDVGOPRRTKAMLKSLKAGGEVIDSVTWHHYL	300					
Db	241	QLGEDYIOLHKLRLKSTFKNAKLYGSDVGOPRRTKAMLKSLKAGGEVIDSVTWHHYL	300					
Qy	301	NGRTATRBDFLPNDVUDIFISSVQKFQVTESTRGKKWNLGESESSAYSGAPLSDTA	360					
Db	301	NGRTATRBDFLPNDVUDIFISSVQKFQVTESTRGKKWNLGESESSAYSGAPLSDTA	360					
Qy	361	AGFMWIDKLGLSARMGIEVWRQVFGAGNTHLUDNFDPWMSLUEKUJTKUM	420					
Db	361	AGFMWIDKLGLSARMGIEVWRQVFGAGNTHLUDNFDPWMSLUEKUJTKUM	420					
Qy	421	ASVQSKRKURLYLHCTNTDPRYKEGDLTLYAALHNVTKYRLPPSNKQDVKYL	480					
Db	421	ASVQSKRKURLYLHCTNTDPRYKEGDLTLYAALHNVTKYRLPPSNKQDVKYL	480					
Qy	481	RPLGPBPGILLSKSVQLINGLTQNDOTPLRMEKPLRPGSSLG	523					
Db	481	RPLGPBPGILLSKSVQLINGLTQNDOTPLRMEKPLRPGSSLG	523					

Search completed: October 22, 2003, 20:24:22  
 Job time : 88 secs





14 IRKMKKTYINGKREKUDIERLKE-----LKEVREHVLVETEDGTYTAKAEDDEEMMRSKV 66

QY 172 VUYTFANCSGLDIFGLNALLRTADLQNSNAQQLDYCCSSKGYNTSWEGLNRP 226

Db 67 GALKEAKYKFAKES-----KTDL-----SNPR-VLDLCSGMGYNATAALHYNK 109

QY 227 NS-----FL-----KKADIFPINGSOLGBDYLQHLKLKSTP 258

Db 110 NABIDMTEICEBSVLELTIPFLDIPYKEHELIKQVREFLN-KIGEY-----KSDY 159

QY 259 KNAKLYGPDVGPERRKTAKMLKSPKLAGGEVIDSVTWHYLYNGRAT--REDPLNPDVL 316

Db 160 DNINLY--VGDAKFIKSDKYY-----NVVFDASFPKRDPLTYDFL----- 202

QY 317 DIFISVQVKVQFVVESTRPGKVKWLGERTSAGGGAPLSDPTFAAGFMWLKLGKLSARMG 376

Db 203 -----KEIVKRMEDN--GVLI-----SYSSATIPRSALVDCGFVISEKESVGRKG 246

QY 377 IEVVMRQVFFGAGNLYLVDENED-----PLDYLWISLFLFKKGVLGKULMASVQSSKR 429

Db 247 ITLAYVQKPNPKRNINSEDEVIALSVALPTDTELSLKQKIDREERREKKEKLI 306

QY 430 KURVYVJACTNTDNPRYKEGDLTYA- INLHNVTYKVRPY 468

Db 307 KIGKYLSTTKQIKKGNIBELKIQKEDLNSSE-IKKORLKE 347

RESULT 5

QY T12094 beta-fructofuranosidase (EC 3.2.1.26) - fava bean

C;Species: vicia faba (fava bean)

C;Accession: T12094

C;Date: 16-Jul-1999 #Sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

R;Weber, H.; Borziluk, L.; Heim, U.; Buchner, P.; Wobus, U.

Plant Cell 7, 1835-846, 1995

A;Title: Seed coat-associated invertases of Fava bean control both unloading and storage

A;Reference number: 217416; NID:9693423; PMID:8551317

A;Accession: T12094

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-575 <NWB>

A;Experimental source: cv. Frib, seed coat

C;Genetics:

C;Gene: CWN1V1

C;Superfamily: beta-fructofuranosidase

C;Keywords: cell wall; glycoprotein; hydrolase

Query Match 3.9%; Score 111; DB 2; Length 670;

Best Local Similarity 22.4%; Pred. No. 3/4; Mismatches 77; Mismatches 194; Indels 156; Gaps 33; Matches 123; Conservative 77; Mismatches 194; Indels 156; Gaps 33;

QY 51 LVSPSPFSLVSTDIAATDPRFILGSPKLRITLARGISPAY--LRFEGTKTDPFLIFDPK- 107

Db 47 LDPSVSPSLVSGN-GVDTQVLPVCGSDLV-----LSPSQPSHVVGKISSWIDSED 99

QY 108 -----KESTEPRHRSVWQSOVNVQDICKYGSIPPDVEBKURLWPMYOBOLLRLBHYQKCFK 161

Db 100 EYRMDSBTTLQKBEIAWATHLISLQMCB-----PD-----LTPHYLAGL 139

QY 162 NSIVSRSSV-----DVIY-----"FANCS--GLDIFGLNALLRTADLQNSNAQQL 207

Db 140 RVSCRSRSPFISBETFLYKITEFNOALTFCGSISFLCNVVISALKWLRLVPLVK--SEGDSM 196

QY 208 LDYCSSSKGYNIEWBLG-----EPNSPLKKA-DIFIN-----GSOLGEYIQLHKUL 253

Db 197 DQ--TSEGGLNDSWELANSPLRICEHSDKSLVSLADVISTLPSSTSLSGWGSS-VRAAITS 253

QY 254 RKFSTFKVAKLYGPDVGP-----RKTAKMLKSET-KAGGEVIDSVTWHYLYNGRATREDF 310

Db 254 TQFLTFLNAR-----GFLCLSKRHKQHQLAGFDHAQWVICKQPKVHNLQKPLDSSGCTB 307

QY 311 LMDVVDIFPISVQVKVQFVVESTRPGKVKWLGERTSAGGGAPLSDPTFAAGFMWLKLG 370

Db 308 KNP--LRYLYDVTAVLFLKRMESLSEQRIBYGRDFLQAPLQPLMDNQLEAQTYETFE-- 362

QY 371 LSARMGIEBVVMQFVFFGAGNLYLVDENEDPLDYLWISLFLKKGVLGKULM----- 420

Db 228 KUPIHSAKRTGWMBCPPFPVPSLLEGKQNLD--LSMMGMMNNVHVKLQNSLIDTRVYKTTG 285

QY 90 AVLRL-----FGGTRKDF-----LIFDPKKESTFERSYW---QSQVNO 124

Db 286 TYLQNQDKYIPDKTSEDCGWGLRYDQYNSFYASKSFDFPTK---NRRITWGANESETKE 341

QY 125 DICKYG-----SIPPDV-----BEKLEWPMQEQQLR---EHQKPKPKNSTYSSV 170

Db 342 DDVVKKGWQGQALPRTWLDSSRQLR-OWPYERBLRNRKGQEMKRLKKGY-----L 396

QY 171 DUYTFANCSGLDIFGLNALLRTADLQNSNAQQLDYCCSKGYNISWELGNEPNFL 230

Db 452 TLA-----SKKLUETTSVFFRVPKAANKHKLIMCSDAKSSSLNRLRBLYKPSFAGFVNVL 505

QY 397 EYKGTQASQADVEVTFSFLDQKAEDPWNWAE---DLCQKQSKVRYG---PQCLL 451

QY 231 KRADIPRINGSOLQEDBYTOL-----HKLL-----RKSTFKNAKUJXGP-----DV 268

QY 524 LPAFSISPFV 533

Db 499 IIS-SVSPFI 507

RESULT 6

T10666 hypothetical protein F6E21.40 - *Arabidopsis thaliana* (mouse-ear cress)

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Feb-2001

C;Accession: T10666

R;Bevan, M.; Leonard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, A; Reference number: Z16533

A;Accession: T10666

A;Molecule type: DNA

A;Residue: 1-670 <BEV>

A;Cross-references: EMBL:AU049914; GSPDB:GN00062; ATSP:F6E21.40

A;Experimental source: Cultivar Columbia; BAC clone F6E21

C;Genetics:

A;Gene: ATSP:F6E221.40

A;Map position: 4

A;Intons: 47/3; 87/1; 123/3; 203/3; 230/2; 255/3; 284/3; 305/1; 335/3; 347/3; 370/3; 355/1

C;Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skbl

Query Match 3.9%; Score 111; DB 2; Length 670;

Best Local Similarity 22.4%; Pred. No. 3/4; Mismatches 77; Mismatches 194; Indels 156; Gaps 33; Matches 123; Conservative 77; Mismatches 194; Indels 156; Gaps 33;

QY 51 LVSPSPFSLVSTDIAATDPRFILGSPKLRITLARGISPAY--LRFEGTKTDPFLIFDPK- 107

Db 47 LDPSVSPSLVSGN-GVDTQVLPVCGSDLV-----LSPSQPSHVVGKISSWIDSED 99

QY 108 -----KESTEPRHRSVWQSOVNVQDICKYGSIPPDVEBKURLWPMYOBOLLRLBHYQKCFK 161

Db 100 EYRMDSBTTLQKBEIAWATHLISLQMCB-----PD-----LTPHYLAGL 139

QY 162 NSIVSRSSV-----DVIY-----"FANCS--GLDIFGLNALLRTADLQNSNAQQL 207

Db 140 RVSCRSRSPFISBETFLYKITEFNOALTFCGSISFLCNVVISALKWLRLVPLVK--SEGDSM 196

QY 208 LDYCSSSKGYNIEWBLG-----EPNSPLKKA-DIFIN-----GSOLGEYIQLHKUL 253

Db 197 DQ--TSEGGLNDSWELANSPLRICEHSDKSLVSLADVISTLPSSTSLSGWGSS-VRAAITS 253

QY 254 RKFSTFKVAKLYGPDVGP-----RKTAKMLKSET-KAGGEVIDSVTWHYLYNGRATREDF 310

Db 254 TQFLTFLNAR-----GFLCLSKRHKQHQLAGFDHAQWVICKQPKVHNLQKPLDSSGCTB 307

QY 311 LMDVVDIFPISVQVKVQFVVESTRPGKVKWLGERTSAGGGAPLSDPTFAAGFMWLKLG 370

Db 308 KNP--LRYLYDVTAVLFLKRMESLSEQRIBYGRDFLQAPLQPLMDNQLEAQTYETFE-- 362

QY 371 LSARMGIEBVVMQFVFFGAGNLYLVDENEDPLDYLWISLFLKKGVLGKULM----- 420

Db 363 -----RDSKVIQOQ--RAVEKALVDR---VPEKEASL-----TIVLWVAGGRGPLV 406

QY 421 -ASVOGSKR--RKLRVYHCTNTDNPRYKEGDLTYA- INLHNVTYKVRPY 468

Db 407 RSLQAMABETDRKVKV--AWEKNEP-----AVTTLHLVLRMEGWEDVUTIISCDM 455

QY 469 PSN--KQDVKYLPIGPPIGHAKISVQVQGLG-----TLMQVDDQTPPLM--EKLRPGSISL 523

Db 456 RPNNAPOBQADLIVSELGLSGF-----DNELSPECIDGAQFQFLKP-DGIS 498

QY 524 LPAFSISPFV 533

Db 499 IIS-SVSPFI 507

RESULT 7

S32961 hypothetical protein YBR259w - yeast (*Saccharomyces cerevisiae*)

N;Alternate name: hypothetical protein YBR1727

C;Species: *Saccharomyces cerevisiae*  
 C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Apr-2002  
 C;Accession: S32961; S46540;  
 R;Doignon, F.; Biteau, N.; Crouzet, M.; Aigle, M.  
 Yeast 9, 189-199, 1993  
 A;Title: The complete sequence of a 19,482 bp segment located on the right arm of chromo-  
 A;Reference number: S29348; MVID:93220397; PMID:465606  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-688 <DOI>  
 A;Cross-references: EMBL:Z36128; NID:9536684; PID:CAA85222.1; PID:9536685; MIPS:YBR259w  
 R;Aigle, M.; Bacle, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.  
 A;Genetics:  
 A;Cross-references: SGD:S000463  
 A;Map position: 2R  
 A;Molecule type: DNA  
 A;Residues: 1-688 <AIG>  
 A;Cross-references: EMBL:Z36128; NID:9536684; PID:CAA85222.1; PID:9536685; MIPS:YBR259w  
 C;Genetics:  
 A;Cross-references: SGD:S000463  
 A;Map position: 2R  
 C;Superfamily: *Saccharomyces cerevisiae* hypothetical protein YBR259w  
 Query Match 3.9%; Score 111; DB 2; Length 688;  
 Best Local Similarity 22.5%; Pred. No. 3.6; Matches 67; Conservative 45; Mismatches 94; Indels 92; Gaps 16;  
 QY 126 ICKYKSSIPPDVVEKURLEWLPYQOBQALRERHOKKERNYSTYRSSS-----VDSLVT 175  
 164 MAEYSSWKWSDDKRQFLQFMYERPRMKKCEIVKVKFENFDLQSKSPLKSLIIPWKEIVV 223  
 QY 176 FANCSGLDULFGNALRITADLQHNSN-----AQQLD-----YCSSSKG----- 216  
 Db 224 -ANC- -IDAFTGEBQVRIDGABLWITPSKRNLYFSSISSAVURLNDIQNMFSARFPYGBEALV 280  
 QY 217 -----NISHELGNEPNSPELKK-----DIF-----INSQLG-----EDVYQHKLK----- 255  
 Db 281 QDFAHIRSLKWDNSDKRVSILRFLAIDNMDPFPYFKEQVDTKADGJFFPLRURKPKFEHN 340  
 QY 256 -----SITFKN--AKLVGPDVGSOPRKTAQMKLKFSLKAGGEV-----IDS 293  
 Db 341 DVKDFHIQVQIKLUNSKOFKNVNSTLMTSSKQSDRKRISHMNSSIIDDGNKIGMHWSPIDE- 399  
 QY 294 TWHRYTLNG-----RPAATDFLNLDPDVLIFPISVQKVFWESTR---PGKK 338  
 Db 400 -YSHFEDNDSPLWRDKVPKYKTFQCTPTPDA5IFDS--HKVIAIISLLRYVPEKR 454  
 RESULT 8  
 T49648 hypothetical protein B8B20.20 [imported] - *Neurospora crassa*  
 C;Species: *Neurospora crassa*  
 C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C;Accession: T49648  
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, A.;Reference number: 225022  
 A;Accession: T49648  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-2298 <SCH>  
 A;Cross-references: EMBL:AL355933; GSPDB:GN00116; NCSP:B8B20.20  
 A;Experimental source: BAC clone B8B20; strain OR74A  
 C;Genetic:  
 A;Gene: NCSP:B8B20.20  
 A;Map position: 6  
 A;Introns: 426/3  
 Query Match 3.9%; Score 111; DB 2; Length 2298;  
 Best Local Similarity 19.3%; Pred. No. 22; Matches 114; Conservative 79; Mismatches 190; Indels 208; Gaps 28;

QY 78 PKURTLAGLSPA-----YLRFGCTKDPFLIPPKKESTPEERSYWSQV-NODIC 127  
 Db 1447 PVDVLERLITPSNAKEACLTINRANNQNLARLUVSNGEGASFRPEITWNVNEQD 1506  
 QY 128 KYSS1PPDVVEKUR-----LEWLPYQOBQALRERHOKKERNYSTYRSSSDVLYTFAN 178  
 1507 QMSAESAEDIEQOBRALSAENNSIDAAMRELEITON-----KATAUDILHTSAR 1555  
 QY 179 CSGIDLIGLNL-----LRTADLO-----VNSNQOLLILYQCSKQYNI 218  
 Db 1556 AS-LDVLKQAKTLEATVTLNTVOLOKMCTIHFGS-PGFDNGILNVAL----DTHAFL 1609  
 219 SW-BLGNEBPNFSKKAIFINGSQLGSDYIQLKURKSTFGQAKLYGPDVQPRKTA 277  
 QY 1610 GWETTSSEBQYSSNSSADIDPQLBDAILLQEKLTKEFPRM-----RELLAL 1659  
 Db 278 MLIKSPFLKAGGEV-D-SVWHHYTLNGRATDFFDLPNPDVLDIFISSVQKVROVESTPG 336  
 Db 1660 PLKAITTFRGKOTEQACTEKTWLAQKAA-----FIO-----FIO-----FIO----- 1706  
 QY 337 K-----A-----KWLGEHNSAYVGGAPLISDTPAAG-----F-----F-----F----- 368  
 Db 1707 KYGLFPDMPKNNSGPERKWL-----PLPIATLUNKNVDFDKDIEINTILSLWQS 1755  
 QY 369 LGUSAR-MGIEVUMRQVPGAGNYH-----VDEIFPDLQWMSL-LPKKLGS-----TKVU--- 419  
 Db 1756 IIKMRFLGQYETVLAELVQORGLPFLAEDVSAGMTPDYNTIHDLFSRAHYMRKALGG 1815  
 QY 420 -----MASVQGSK-----RKURVYIHCNTDNPRYKEGDLTYIAINAHNVTKL 464  
 Db 1816 ATTPAPGVTIS-SATAGSSAQSIIRQREFSH----- 1847  
 QY 465 RLPLPFSNQVQDKYTLRPLG-----PHGLIISKSVQVQNGIITLKWMD- 504  
 Db 1848 TLQLMATNIKQDIFLRLSLAADPTASSTEERDYMATHGLS-LIKSHGVGIVVVDSP 1906  
 QY 505 -----DQTLPPIMEKPKRPG-----SISGLPAPSYR-FVTRNKA 540  
 Db 1907 FLTPSDSYSPPLQDPPQHQTAGIMAYGVRLSKDVKVPAASQOLPWLFNNPKVA 1957  
 RESULT 9  
 E91031 probable outer membrane protein ECs3221 [imported] - *Escherichia coli* (strain O157:H7, *E. coli* O157:H7) - *Escherichia coli* (strain O157:H7, *E. coli* O157:H7)  
 C;Species: *Escherichia coli*  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C;Accession: E91031  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene  
 A;Reference number: A99629; MVID:21156231; PMID:11258796  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-879 <HAY>  
 A;Cross-references: GB:BA000007; PID:BAB26644.1; PID:gi|3362691; GSPDB:GN00154  
 A;Experimental source: strain O157:H7; sub-strain RIMD 0509952  
 C;Genetic:  
 A;Gene: ECs3221  
 Query Match 3.9%; Score 109.5; DB 2; Length 879;  
 Best Local Similarity 20.1%; Pred. No. 6.7; Matches 130; Conservative 68; Mismatches 208; Indels 241; Gaps 33;  
 QY 52 VSPRLPSVTDIANLATAIDPFLTIGSPKURTLARGSLPSAYLRFGCTKDPFLIPPKKEST 111  
 Db 20 MSGSYVNAWANEIOPDSRFLELKGDTK-----LKRPSGCGYVERG-KYNLQVOLNKGPLT 76  
 QY 112 FEENSYWQSVQNDICKYGS1PPDVVEKU-----LEWLPYQOBQALRERHOKKERN 163  
 Db 77 EBDYIYWAESENDASKTVACTTLEBLVAFGLKEDVAKNLQWHDGKELKPGOLE----- 130

QY 364 TYSRSVTDVLFVAFNGSGDLIFGL-NALNIRTDQGNSN-----AQIJDYC---- 211  
 Db. 131 -----GIDIK--AddSOSALVTSRQAYLEVDINWDPSPRWDGDCISGLIADYSTAQT 182  
 QY 212 -----SSKGNTI-SSEL-GNPNSTI--KKADIFINGSO----- 241  
 Db. 183 EHHENGSSDRSNSRISGNTVYGVNIGARLARAWQTDVLSKSENDDVINGDQTKQWMSR 242.  
 QY 242 -----LGBDYLQHLKURKSTP----- 268  
 Db. 243 YIAWRLPSLAKAKGGEDY-----LNSDTEFGFNVGGS1STDQMLPPNLRGAPDI 296  
 QY 269 GQPRRTAKMKSPLKAGGEGV----- 312  
 Db. 297 SGVAHTYAKTVSOL--GRVYIETQVPGAFPRRIQDGV-----SGTHIRREBQN 346  
 QY 313 PDVLDIFISSVQVKVQVNESTRPGK--KVNL----- 345  
 Db. 347 GVOQEDINTASMPP--LTRPGQRYKUMGRPQEWGHVVEGGFSGGASWGLANG 402  
 QY 346 SAYGSGAPLSD-----TEAA-----GEWFLDKU-----GLSAR 374  
 Db. 403 SLYGGA--LADSHYOSAALGVGVRDLSVEGAFAVDFTHSHTRLDKETAYGKSLDONSFR 459  
 QY 375 MGI----EVURQVFGAGHILVNEFDLDPWVSLFELKJQV--KULMASVQGS 426  
 Db. 460 LSYSKPQFDELSRSVTAG--YRFSEENFMWSY-LDASDSEMTRTGNDKEMVYATYNO 515  
 QY 427 KERKURKQVYLHCTNDPQYKEGDLTYAI-----NLUHNVTK----- 462  
 Db. 516 NFRDAGSVVWVATRATWDFDQTCINWMSHFWLGSIRKMSMTGIRKIEYDQADK 575  
 QY 463 --YLRLPPYFPEKQVKYLUPLGPGLLSVQNGLTLWMDPQ 507  
 Db. 576 GYVVISLSPWQDPSSTSY--NGNYGSGDSSQVG--YFSRVDDDT 616  
 RESULT 10  
 S7760 Hypothetical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)  
 C;Species: *Synechocystis* sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C;Accession: S74760  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, A.;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. N 1617  
 A;Reference number: S74322; MUID:97061201; PMID:8905231.  
 A;Cross-references: EMBL:D90901; GB:AB001339; NID:91651897; PIDN:BA16911.1; PID:di01764  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residue: 1-411 <KANJI>  
 A;Cross-references: EMBL:D90901; GB:AB001339; NID:91651897; PIDN:BA16911.1; PID:di01764  
 QY 305 ATREBLAPD-----LDIRISSVQKVQVQFESTRPGKKWVIGETSSAYGGAPLSDTEA 360  
 Db. 312 -----NPKLUDLHDDLSSTIVATL-----KSPRIGNANT-----GTCQSLSTLK 351  
 QY 361 AGFMWIDKUGLSA---RNGIEVNMQVFFGAGYHLYDENFDPDLYWLSLFLKLV 415  
 Db. 352 FAMTRDEGSSSMQILENTVSIAMVYGRAN-HVLD-----WEPVIRF-OC 400  
 QY 416 TKVLMASV 423  
 Db. 401 LKSLUHQI 408  
 RESULT 11  
 D87541 beta-xylanidase [imported] - *Caulobacter crescentus*  
 C;Species: *Caulobacter crescentus*  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C;Accession: D87541  
 R;Nierman, W.C.; Feldblyum, T.V.; Päusen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Laub, M.T.; deBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolodkin, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 336-4141, 2001  
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A;Reference number: A87249; MUID:21175698; PMID:11256647  
 A;Accession: D87541  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-500 <STO>  
 A;Cross-references: GB:AE005673; NID:913423886; PIDN:AAK24328.1; GSPDB:GN00148  
 C;Genetics:  
 C;Gene: GC2357  
 Query Match 3.0%; Score 107.5; DB 2; Length 500;  
 Best Local Similarity 25.7%; Pred. No. 4; Mismatches 80; Indels 47; Gaps 15;  
 Matches 56; Conservative 35; Mismatches 80; Indels 47; Gaps 15;  
 QY 165 YRSSVTDVLFVAFNGSGDLIFGK--ALLRTRD--LW-NSSNAOL---LLD-- 209  
 Db. 81 YDWTKIDOLYDALLAKGKIFKELGFTPEAMKTSQTIYWKGNITSHPKLGPWRDLDIF 140  
 QY 210 -YSSKGNTI--SW-ELGNEEN-SEKKKAIIFNSQLGQDPIYQHKLURKSTFK 260  
 Db. 141 VHLRPARYGVETEVRHPEFEWTNEENLNGWKEAD-----QAYFELYDV---TARA 188  
 QY 261 AKLYGPD--VGQPRRTAKMKSFL--KAGGEVTDVSWHHYLYNG---RTATRDEL 311  
 Db. 189 IKAIQPSLRVYGGATAGAAWVPELARVKKSGSAVDFVTHYGVGDDGFLDEKSYQDTKL 248  
 QY 312 NPVDLIFISSVQVKVQVNE-STRPGKKWVLGESSAY 348  
 Db. 249 SPSP-DAVGDVRRVREQIEBASAFPGLPLYFTENSTSY 285  
 RESULT 12  
 F8875 probable fimbrial usher 23600 [imported] - *Escherichia coli* (strain 0157:H7, substrain E) C;Species: *Escherichia coli*  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C;Accession: F8875  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, R.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouidis, K.; Apodaca, L.; Nature, 409, 559-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: F8875  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residue: 1-879 <STO>  
 A;Cross-references: GB:AE005174; NID:912516702; PIDN:AAG57466.1; GSPDB:GN00145; UMGD:Z306  
 A;Experimental source: strain 0157:H7, substrain ED1933  
 C;Genetics:

Query Match 3.7%; Score 106; DB 2; Length 879;  
 Best Local Similarity 20.5%; Pred. No. 12; Mismatches 201; Indels 234; Gaps 34;

Matches 129; Conservative 64; Mismatches 201; Indels 234; Gaps 34;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6; Mismatches 149; Indels 86; Gaps 15;

Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

A; Cross-references: EMBL:AB000269; NID:93341860; PID:BA31857.1; PID:g3341861  
 C; Genetics:  
 A; Gene: ssm4; SPAC27D7.13c  
 A; Map position: 1

A; Status: preliminary; translated from GB/EMBL/DBJ

RESULT 13

T38446 Ssm4-associated protein Ssm4 - fission yeast (Schizosaccharomyces pombe)

C; Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000

C; Accession: T38446; T00012; Barrell, B.G.; Rajandream, M.A.

R; McDouall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, December 1997

A; Reference number: Z21793

A; Accession: T38446

A; Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: DNA

A; Residues: 1-788 <MCR>

A; Cross-references: EMBL:X06601; PID:94903; PID:CAA29820.1; PID:94904

R; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.

submitted to the EMBL Data Library, February 1998

A; Reference number: Z21910

A; Accession: T40496

A; Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: DNA

A; Residues: 1-788 <WOO>

A; Cross-references: EMBL:AL021730; PID:CAA16823.1; GSPDB:GN00067; SPDB:SPBC4C3..02C

A; Experimental source: Strain 972h; cosmid C27D7

R; Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, March 1999

A; Reference number: Z21928

A; Accession: T40522

A; Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: DNA

A; Residues: 604-788 <SRE>

A; Cross-references: EMBL:AL03555; PID:CAA8600.1; GSPDB:GN00067; SPDB:SPBC4C05..01C

A; Experimental source: Strain 972h; cosmid C405

A; Status: translated from GB/EMBL/DBJ

A; Molecule type: DNA

A; Residues: 1-670 <YAM>



GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 22, 2003, 20:21:10 ; Search time 25 Seconds  
 (without alignments)

982.135 Million cell updates/sec

Title: US-09-759-207-2  
 Perfect score: 2842

Sequence: 1 MLLRSKCALPPPLMLLULGP.....LPAFSYSEFVNKAACI 543

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112.5	4.0	356	1 Y670_METU	058084 methanococcus
2	111.1	3.9	688	1 Y89F_YEAST	P38654 schizosaccharomyces
3	105.5	3.7	1031	1 TERT_EUPAE	000939 eu pliotes ae
4	105.0	3.7	670	1 SSM4_SCHPO	042667 schizosaccharomyces
5	104.5	3.7	646	1 THIC_VIBRA	Q87K10 vibrio para
6	104.5	3.7	788	1 PUR2_SCHPO	P20772 s bifunction
7	104.3	3.7	432	1 PURA_AQUAB	067221 aquifex aeo
8	104.1	3.7	716	1 RRP2_TAXIT	091742 influenza a
9	103.6	3.6	796	1 PHK_CLOAB	0977e3 clostridium
10	101.0	3.6	644	1 SYM_CLOAB	0978e3 clostridium
11	100.5	3.5	805	1 GYRB_CHLPN	0928f3 chlamydia p
12	99.5	3.5	837	1 KXNZ_CLOTM	P10478 clostridium
13	99.3	3.5	4594	1 DYHC_HUMAN	Q14204 homo sapien
14	99.3	3.5	4644	1 DYHC_RAT	P38650 rattus norvegicus
15	98.5	3.5	629	1 T3MR_HAEIN	P71366 haemophilus
16	98.5	3.5	654	1 NUSM_RHIST	P50367 rhizopus st
17	98.5	3.5	746	1 RHTA_RHIME	P09236 rhizobium m
18	98.5	3.4	716	1 RRP2_IAV17	P31343 influenza a
19	97.5	3.4	454	1 YUA6_CABEL	P52715 ceborhabdium
20	97.5	3.4	804	1 GYRB_CHLPN	084193 chlamydia t
21	97.5	3.4	1314	1 SS22_YEAST	P25390 saccharomyces
22	96.5	3.4	595	1 THIC_BACHD	092014 bacillus haemolyticus
23	96.5	3.4	327	1 XWNA_ASCAC	059859 aspergillus
24	96.5	3.4	557	1 COX1_NEUCR	P03945 neurospora
25	96.5	3.4	716	1 RRP2_TAZ1	P13175 influenza a
26	96.5	3.4	1044	1 ITAV_MOUSE	P43406 mus musculus
27	95.5	3.4	358	1 VALL_BCTV	P14991 beet curly top virus
28	95.5	3.4	620	1 HEMA_MEASY	P28081 measles virus
29	95.5	3.4	525	1 BACMA_BACLI	068006 bovine calicivirus
30	95.5	3.3	716	1 RRP2_IAMOR	P13170 influenza a
31	95.3	3.3	772	1 LPIG_DROME	P11997 drosophila
32	95.3	3.3	2214	1 SORBL_HUMAN	Q92673 horse
33	95.3	3.3	4644	1 DYHC_MOUSE	Q9jhua4 mus musculus

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_41,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	3.3	592	1 INV2_DAUCA	Q39692 daucus carota
2	94.5	3.3	645	1 SYM_CLOPE	Q8dhg1 clostridium
3	94	3.3	437	1 BRF1_XENLA	P55615 xenopus laevis
4	94	3.3	804	1 GYRB_CHLMD	Q5613 chlamydia m
5	93.5	3.3	657	1 GRAD_TREPA	P03062 treponema pallidum
6	93.5	3.3	766	1 GAF1_SCHPO	P33277 schizosaccharomyces pombe
7	93.5	3.3	1787	1 UVRA_CHLMD	Q9pk60 chlamydia m
8	93	3.3	449	1 A4ML_MOUSE	Q9jkct7 mus musculus
9	93	3.3	726	1 CATE_BOV1	P13021 escherichia coli
10	93	3.3	2733	1 RRPB_CVMAS	P16342 murine coro
11	92.5	3.3	455	1 MURP_BUCA1	P37315 buchnera apiculata
12	92.5	3.3	738	1 YASS_SCHPO	Q10145 schizosaccharomyces pombe

Scoring table: BiOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announcements> or send an email to license@isb-sib.ch).

Db 846 DWGIGSINKTKTALMPNINLDRIGLCTNLNNMOTKASMLKKRKGSPF----MINI 899  
 Qy 294 TWHYHYNGRTATRBDPLNFDLPISSVQKPVQVVESTRGKVKWLGTSAYGGAP 353  
 Db 900 T----RKTITPDRANKTLKLPISSGKYKNCQAKYKDFRKNTAMSSMIDLVSK 955  
 Qy 354 LUDTRPANGPMWMDKLGLSARMGVIEVWVQVPGAGNHLVLDNFDPLDPLDYLWLSL---- 409  
 Db 956 IIVSVTRAPPKYL----VCNICKTIFGEEHY----PDFLFLSTKHF 994  
 Qy 410 ---PKKUVTGKTVLMASTVOGSKRKRLR 432  
 Db 995 BIFSTKQVIFPNRVCW---ILKAKERKLK 1019  
 RESULT 4  
 SSM4\_SCHPO STANDARD; PRT; 670 AA.  
 ID SSM4\_SCHPO  
 AC 042667  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-PBS-2003 (Rel. 41, Last annotation update)  
 DB Microtubule associated protein Ssm4.  
 GN SSM4 OR SPAC27D.13C  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Bokai-Yo, Pungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP  
 RX MEDLINE=97311255; PubMed=9167972;  
 RA Yamamoto A., Watanabe Y., Yamamoto M.,  
 RT "Microtubule-associated coiled-coil protein Ssm4 is involved in the  
 meiotic development in fission yeast.";  
 RT Genes Cells 2:155-166(1997).  
 RL  
 RN  
 SEQUENCE FROM N.A.  
 RP  
 RX STRAIN=972;  
 RA MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., RaJandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., RaJandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle B.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbitts E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Wolkaert G., Aert R., Robben J., Gromyko B.,  
 RA Weltjens I., Vandervelde B., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzsche C., Holzer B., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Delaure V., Mottier S.,  
 RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochelet M., Gaillard C., Tellada V.A., Gazon A., Thode G.,  
 RA Daga R.M., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
 RA Cerritoni L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RA Nature 415:871-880(2002).  
 CC -> FUNCTION: Binds to nuclear microtubules with the effect of either  
 CC modifying their structure or function. This then promotes meiotic  
 CC nuclear division.  
 CC -> SUBCELLULAR LOCATION: Mitotic spindle.  
 CC -> SIMILARITY: Contains 1 CAP-Gly domain.  
 RESULT 5  
 THIC\_VIBPA STANDARD; PRT; 646 AA.  
 ID THIC\_VIBPA  
 AC 087KFO  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB Thiamine biosynthesis protein thiC.  
 GN THIC OR VF3027.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OC NCBI\_TaxID=670;  
 SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;







ID	GYRB	CHLBN	STANDARD;	PRT;	805 AA.
AC	_GYRB_CHLBN				
Q9ZBR3 ; Q9JQ44 ;					
DT	30-MAY-000 (Rel. 39, created)				
DT	30-MAY-2000 (Rel. 39, last sequence update)				
DT	16-OCT-2001 (Rel. 40, last annotation update)				
DB	DNA gyrase subunit B (EC 5.99.1.3.)				
GN	gyrb OR CPN025 OR CP0484				
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae)				
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila				
OX	NCB1_TAXID=83558;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CML029;				
RX	MEDLINE=92060606; PubMed=10192386;				
RA	Kalmán S., Mitchell W., Martínez R., Lammel C., Fan J., Hyman R. W.,				
RA	Olinger L., Gilwood J., Davis R. W., Stephens R. S.;				
RT	"Comparative genomes of <i>Chlamydia pneumoniae</i> and <i>C. trachomatis</i> .";				
RL	Nat. Genet. 21:385-389 (1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AR39;				
RX	MEDLINE=20150255; PubMed=10684935;				
RA	Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,				
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,				
RA	Linher K., Wedman J., Khoruts H., Craven B., Bowman C., Dodson R.,				
RA	Gwin M., Nelson W., Debby R., Kolonay J., McClarty G., Salzberg S.L.,				
RA	Eisen J., Fraser C.M.,				
RT	"Genome sequences of <i>Chlamydia trachomatis</i> MoPn and <i>Chlamydia pneumoniae</i> AR39.";				
RL	Nucleic Acids Res. 28:1397-1406 (2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=J138;				
RX	MEDLINE=20330449; PubMed=10871362;				
RA	Shimai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,				
RA	Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.;				
RT	"Comparison of whole genome sequences of <i>Chlamydia pneumoniae</i> J138 from Japan and CML029 from USA.";				
RT	Nucleic Acids Res. 28:2311-2314 (2000).				
CC	-1- FUNCTION: DNA GYRASE NECESSARILY SUPERCOILS CLOSED CIRCULAR DOUBLE STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE				

INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED DNA RINGS, INCLUDING CATEINANES AND KNOTTED RINGS. CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining of double-stranded DNA.

-1- SUBUNIT: MADE UP OF TWO CHAINS: THE A CHAIN IS RESPONSIBLE FOR DNA BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE ENZYME FORMS AN A2B2 TETRAMER.

-1- SIMILARITY: Belongs to the type II\_topoisomerase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB001612; AAD18424; 1; DR EMBL; AE002110; AAF8314; 1; DR EMBL; AP002546; BAA8485; 1; DR PIR; C862525; C86525; DR PIR; H72098; H72098; DR HSSP; P05982; IAU6; DR PHC1-2DRG8; Q9Z8R3; DR TIGR; CP0484; DR InterPro; IPR003594; ATPbind, Atppase, InterPro; IPR002288; DNA\_gyraseB\_C, InterPro; IPR00565; DNA\_gyrB, InterPro; IPR01241; DNA\_topoisom. InterPro; IPR006171; Toprim dom. DR Pfam; PF00204; DNA\_gyraseB; 1, DR Pfam; PF00986; DNA\_gyraseB\_C; 1, DR Pfam; PF02518; HATPase\_c; 1, DR Pfam; PF01751; Toprin; 1, DR PRINTS; PRO0416; TP2FFAMIN, DR Prodom; P014963; DNA\_gyraseB; 1, DR SMART; SM00387; HATpase\_c; 1, DR SMART; SM00433; TOP2c; 1, DR TIGR4MS; TIGR0159; GYRB; 1, DR PROSITE; PS01017; TOPOISOMERASE\_II; 1, DR Topoisomerase; Isomerase; ATP-binding; Complete proteome, SQ SERPINE 805 AA; CO82DF4CCB7BCC CRC64;

Query Match 3.5%; Score 100.5; DB 1; Length 805; Best Local Similarity 20.4%; Pred. No. 12; Matches 121; Conservative 71; Mismatches 169; Idents 231; Gaps 32; QY . 48 PLHLVSPS-----FLSFTIDANLATORPFLTILGSKRKTARGISPAYL 92

Db 153 REHVKKERNSTYRSRSSVVDLYTPANGSLGLDLIFGLNALLRTRADIQWNSNSAQQLDYS 212

Db 237 --YLNQNKESLFS---EPTVIGCPRGVDD---GEIEFBAALQWNSGSELVSYAN 284

QY 93 RFGGTKDFLFDPKKESTPERRSYMSQVNQDICKYGSITPPDVEKLRLWPPYQBRQLL 152

Db 213 SKGYNT-SWELGNBPNPLKKADIFNGSQLGEDYIQLHKLURKSTPKNAK-LIGPDVG 269

Db 285 --NFTTROGQGTHLGPSTALTRVIN---TYKANHIA---KNNKLALQEDI- 328

Db 329 ---D-----REGITAVISKVKNPQEGQ--TKQKLGNSDSSVAQQVGBALTI 372

QY 270 OPRKTKAKNOKLESFLKAGBGEVDISVTWHYIYLNGTATREDFLNPDV-----LDI 318

Db 319 FISS-----QWVESTR-PCK-----K 338

Db 373 FREENPOIARMVTDKVFVAQARAKKARELTLSKALDSARLPGSKLIDCLEKOPKECE 432

QY 339 WLGEGTSSAYGCGA-----PLSDPTFAAGEMWLDKGLSARMGIEVVMRQVFF 386

Db 433 MIVVEGCGA-GGSAKQGRDRRQQAIIIRGKLNVEGARLQIFQNOEIGTTIAAGCGI 491

QY 387 GAGSYHVDENFDPLPDWLSLFLPKUNGKVYLM--ASVQSKRKURV--YLHCTN-- 439

Db 492 GADMFNL-----SKURYRI---IMTDADVDSHIRTLLTFFYRNTALI 535

QY 440 -----TDNPVYKEGDITLYAINLNHNTKVLRLPFSKQVYLRPGCPH--GLS 490

Db 536 ENEVVYIAQPLUK-----VSCKDQPRYIUSEKENDSYLM-LGPNESSILF 581

QY 491 KSV--QLINGLTF---MVDOTLPLMKEQLRPGS-----SIGLPAF 527

Db 582 KSPERLGEALESFINVILDVESFINTLKEKAIPFSEBFLEMVYKEGIGGYPL 633

RESULT 12

XYNZ	CLOM	STANDARD	PRT;	837 AA.
ID	XYNZ_CLOM			
AC	PI0478;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DB	Brdo-1,4-beta-D-xylanase Z precursor (EC 3.2.1.8) (Xylanase Z)			
DB	(1,4-beta-D-xylan xylanohydrodrolase Z).			
GN	XYNZ.			
OS	Clostridium thermocellum.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OX	Clostridium.			
RN	[1].			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=NCTB 10682;			
RP	MEODINB89000072; PubMed=3139632;			
RA	Gréchet O., Chebrou M.;C., Begin P.;			
RT	"Nucleotide sequence and deletion analysis of the xylanase gene (xynZ) of Clostridium thermocellum.";			
RT	(xynZ) of Clostridium thermocellum.			
RL	J. Bacteriol. 170:4582-4588(1988).			
RN	[2].			
RP	X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS) OF 515-837.			
RC	STRAIN=NCTB 10682;			
RC	MEODINB89000072; PubMed=7664125;			
RA	Dominguez R., Souchon H., Spinnelli S., Dauter Z., Wilson K.S.,			
RA	Chauvaux S., Beguin P., Alvari P.M.;			
RT	"A common protein fold and similar active site in two distinct families of beta-glycanases."			
RT	Nat. Struct. Biol. 2:569-576(1995).			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic linkages in xyloans.			
CC	-1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS WELL AS IN OTHER C. THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.			
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL HYDROLASES).			
CC	-1- SIMILARITY: Contains 1 xynz-type cellulose-binding (CBD) domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; M24624; AAA23286; 1; DR			
DR	PDB; 1JTF; 31-OCT-01.			
DR	PIR; A31842; A31842.			
DR	PDB; 1XZ7; 29-JAN-96.			
DR	PDB; 1JTF; 27-MAR-02.			
DR	InterPro; IPR0634; CDD-IV.			
DR	InterPro; IPR0505; CDD-5.			
DR	InterPro; IPR02105; Dockerin_1.			
DR	InterPro; IPR02048; EG-hand.			
DR	InterPro; IPR00801; Esterase_put.			

DR InterPro; IPR001000; Glyco\_hydro\_10.  
 DR InterPro; IPR000379; Ser\_estrB\_site.  
 DR Pfam; PF00342; CBM\_6\_1; 1.  
 DR Pfam; PF00044; Dockerin\_1; 2.  
 DR Pfam; PF00755; Esterase\_1.  
 DR Prints; PRO0134; GLYTHORLABB10.  
 DR SMART; SM00606; CBD-TM\_1.  
 DR SMART; SM00633; GLYC\_10\_1.  
 DR PROSITE; PS00018; BPFAND; UNKNOWN\_2.  
 DR PROSITE; PS00448; CLO5\_CELLULOSE; CPT; 2.  
 DR PROSITE; PS0591; GLYCOSYL\_HYDRO\_E10; 1.  
 KW Xylian degradation; Hydrolase; Glycosidase; Repeat; Signal; 3D-structure; 1.  
 FT SIGNAL 28  
 FT CHAIN 29  
 FT ACT SITE 645  
 FT ACT-SITE 645  
 FT DOMAIN 430  
 FT REPEAT 453  
 FT DISULFID 783  
 FT TURN 518  
 FT STRAND 525  
 FT HELIX 528  
 FT TURN 533  
 FT STRAND 535  
 FT HELIX 536  
 FT TURN 539  
 FT TURN 540  
 FT STRAND 540  
 FT HELIX 543  
 FT TURN 555  
 FT HELIX 561  
 FT STRAND 562  
 FT TURN 564  
 FT STRAND 570  
 FT TURN 571  
 FT STRAND 572  
 FT HELIX 573  
 FT TURN 577  
 FT STRAND 589  
 FT HELIX 590  
 FT TURN 592  
 FT STRAND 599  
 FT HELIX 606  
 FT TURN 610  
 FT STRAND 615  
 FT HELIX 633  
 FT TURN 639  
 FT STRAND 644  
 FT TURN 647  
 FT STRAND 649  
 FT TURN 654  
 FT STRAND 654  
 FT HELIX 658  
 FT TURN 665  
 FT HELIX 667  
 FT TURN 680  
 FT STRAND 683  
 FT HELIX 696  
 FT TURN 711  
 FT STRAND 717  
 FT HELIX 723  
 FT TURN 730  
 FT STRAND 745  
 FT TURN 746  
 FT STRAND 749  
 FT TURN 751  
 FT STRAND 762  
 FT HELIX 765  
 FT TURN 787  
 FT STRAND 789  
 FT TURN 794  
 FT STRAND 798  
 FT TURN 799  
 FT STRAND 800  
 FT HELIX 804  
 FT TURN 809  
 FT STRAND 811  
 FT TURN 812  
 FT STRAND 817  
 FT TURN 819  
 FT STRAND 820  
 FT TURN 823  
 FT STRAND 823  
 SEQUENCE 837 AA; 9262 MW; DD4C29F04D12B6CD CRC64;

Query Match 3.5%; Score 99.5; DB:1; Length 837;  
 Best Local Similarity 19.3%; Pred. No: 15; Mismatches 107; Indels 109; Gaps 16;  
 Matches 63; Conservative 48; Mismatches 107; Indels 109; Gaps 16;  
 Qy 145 PFOEQILRS---HYQCKE---KASTYSSSVUYTFRANCGLDILRGNALR 193  
 Db 543 PTYNSTLQRPSMVYCENELKFDALQPROVFDPSKGDLQLAFAERNGHQ---MR 594  
 Qy 194 TADLQNSNSNQQLID-----YCSSKGNTSWEGLNE---PNSF 229  
 Db 595 GHTLWHNQPSWITNGNRDSSLAWMKHITTMKQKIVBWDVANECMDSGNL 654  
 Qy 230 LKKADIFPNSQLDGYIOLHKLKSTPKAKYGP-----DVGQPRRKTAK 277  
 Db 655 --RSSIWRN--VIGQPDYD-----AFYRAYEADPDALLFYNDYNTEDGPKSNAVEN 703  
 Qy 278 MRSKPLKAGGEVBDIVTMWHLNLTARFEDFNPDV---LDIFISSVQKQVQTES 332  
 Db 704 MIKS-MKERSGVPDQGFGQCHFINGMSPEYLASIDQNTKIRYABGIVVUSFTETIIRIPQ 762  
 Qy 333 TRPG-----KRYWLGTSSSAVCGGAPILSDTAAFGRMW-LDKLGLSARGMT 377  
 Db 763 ENPAPFQVQDNYKELMKICLAPN-----CMTFV--MMGFIDKX----- 801  
 Qy 378 EVMMQVQFQAGNYILVDEDFDPLPDY 404  
 Db 802 -TWIQTFFPSYGNPLIYDSYNNPKDVA 827

RESULT 13  
 DYHC\_HUMAN  
 ID DYHC\_HUMAN STANDARD; PRT; 4594 AA.  
 AC Q14204; 092814; Q914G5;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE Dynen heavy chain, cytosolic (DYHC). (Cytoplasmic dynein heavy chain  
 DE 1) (DYHC1) (Fragment).  
 GN DNCH1 OR DNECL OR KIAA0325.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Ohara O., Nagase T., Kikuno R., Yamakawa H., Nomura N.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 'SEQUENCE OF 1078-1974 FROM N.A.  
 RX MEDLINE=6234671; PubMed=8666668;  
 RA Vaiberg E.A., Griswom P.M., McIntosh J.R.,  
 RT "Mammalian cells express three distinct dynein heavy chains that are  
 localized to different cytoplasmic organelles.";  
 RT J. Cell Biol. 133:831-842 (1996).  
 RN [3]  
 SEQUENCE OF 1932-1972 FROM N.A.  
 RX MEDLINE=404467; PubMed=8227145;  
 RA Vaiborg E.A., Koome M.P., McIntosh J.R.,  
 RA "Cyttoplasmic dynein plays a role in mammalian mitotic spindle  
 formation.";  
 RT J. Cell Biol. 123:849-858 (1993).  
 RN [4]  
 SEQUENCE OF 2508-4594 FROM N.A.  
 RX MEDLINE=7349884; PubMed=9205841;  
 RA Nagase T., Ishitaka K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro.,"  
 RT DNA Res. 4:141-150 (1997).

CC -1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules. CC Dynein has ATPase activity; the force-producing power stroke is thought to occur on release of ADP. CC SUBUNIT: Consists of at least two heavy chains and a number of intermediate and light chains. CC SUBCELLULAR LOCATION: Cytoplasmic. CC (which binds cargo and interacts with other dynein components), and the head or motor domain. The motor contains six tandemly-linked AAA domains in the head, which form a ring. A stalk-like structure (formed by two of the coiled coil domains) protrudes between AAA 4 and AAA 5 and terminates in a microtubule-binding site. A seventh domain may also contribute to this ring; it is not clear whether the N-terminus or the C-terminus forms this extra domain. There are four well-conserved and two non-conserved ATPase sites, one per AAA domain. Probably only one of those (within AAA 1) actually hydrolyzes ATP, the others may serve a regulatory function. -1- SIMILARITY: Belongs to the dynein heavy chain family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB002323; BAA20783; 2. CC EMBL: US3530; BAB0927; 1. CC EMBL: L23958; AAA16055; 1. CC PIR: A49019; A49019. CC PR: G02529; G02529. CC GeneID: HGNC:2961; DNC1. CC MM: 600112. CC GO: GO:0005868; C:cytoplasmic dynein complex; NAS. CC GO: GO:0008567; P:dynein ATPase activity; NAS. CC GO: GO:0007052; P:mitotic spindle assembly; NAS. CC InterPro: IPR005613; AIP3. CC InterPro: IPR004273; MND; dynein heavy. CC Pfam: PF0028; Dynein heavy; 1. CC Pfam: PF03915; AIP3; 1. CC SWISS-PROT: AAA; 4. CC Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil. CC KOW

FT NON\_TER 1 DOMAIN 1 1815 STEM (BY SIMILARITY). CC KOW

FT DOMAIN 1816 2047 AAA 1 (BY SIMILARITY). CC KOW

FT DOMAIN 2128 2400 AAA 2 (BY SIMILARITY). CC KOW

FT DOMAIN 2504 2753 AAA 3 (BY SIMILARITY). CC KOW

FT DOMAIN 2847 3116 AAA 4 (BY SIMILARITY). CC KOW

FT DOMAIN 3137 3448 STALK (BY SIMILARITY). CC KOW

FT DOMAIN 3501 3730 AAA 5 (BY SIMILARITY). CC KOW

FT DOMAIN 3953 4169 AAA 6 (BY SIMILARITY). CC KOW

FT DOMAIN 129 150 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 403 426 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 491 514 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 1119 1200 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 1305 1323 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 3137 3223 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 3344 3448 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 3685 3748 COILED COIL (POTENTIAL). CC KOW

FT DOMAIN 1854 1861 ATP (POTENTIAL). CC KOW

FT DOMAIN 2172 2179 ATP (POTENTIAL). CC KOW

FT DOMAIN 2520 2553 ATP (POTENTIAL). CC KOW

FT DOMAIN 2885 2892 ATP (POTENTIAL). CC KOW

FT DOMAIN 3126 3127 LH--> SD (IN REF. 2). CC KOW

FT DOMAIN 1889 1889 M--> R (IN REF. 2). CC KOW

FT DOMAIN 1973 1973 R--> N (IN REF. 2). CC KOW

FT DOMAIN 3977 3977 H--> O (IN REF. 4). CC KOW

SEQUENCE 4594 AA; 527282 MW; B6P8DAFF1730169 CRC64;

Query Match 3.5%; Score 99; DB:1; Length 4594; Best Local Similarity 19.7%; Pred. No. 2e+02; Matches 113; Conservative 82; Mismatches 172; Indels 208; Gaps 27; Matches 113; Conservative 82; Mismatches 172; Indels 208; Gaps 27; RT 1291 :|DQMKEQPVNSVQPKRQNLNDLNLQKSF----PARLQVASYERVQRURJKYKIN 1345 RT 101 FLIPDPKKESTFEERSYNSQVNO----DICKYGSIPPV---- 136 RT 1346 MLVIEDEKSKA-LKQR-HWQQLMQRHLVVWVVSUTLGQIWVVDLQKNAIVKQVLLVAQ 1403 RT 137 :|EE--KURLEWPMQYQQLLRLAHEHQKPKRNTYSRSSVVDLYTFANCSGLDILGILN 189 RT 1404 EMALBEFLKIQKREVNWTYDLYVW-NYONCR---- 1439 RT 190 ALLRTADLQNSHSSAQQLLUDYCSESKGYN----SWEGLBNSFLKKADIFIN---- 238 RT 1440 DLFNKVKEITNSVAMKLSPP----YKVEEDALSWE----DKUNRIMALFDWIDVQRW 1492 RT 239 :|GS----QGABDYTOIHLKLKAKTSPRNKAQYGPDVQGQPRKT 275 RT 1493 VLGEGIFTGESSADIKHLLPVETQRQISTEFLALMKKVSPLVMDVNLNIGQVRSRLU 1552 RT 276 AKMKSFLAGGEGV----DSVTWVHYYLQGRATREDPFLNPDPDIF----TSSVQYFQ 328 RT 389 GNYVHLVUDENFDPLDQWYSLILFKLKGUVGKVLMSVQGSKRKRKURVHLCTNTDPRYKEG 448 RT 1638 :|VSIEHPKINNEWLTIVKEEM----RVTIQLAKLAVSIVIEWBIFGKATSBIPNTY-- 1686 RT 329 WESTRPGKXWVNLGETSSAYGGAPLSDTPAAGPMWMDLKGLSARMGTEVWVQRQVFGA 388 RT 1603 :|KMPAGVSSI----ILBNDNSV----VLGISSEREGEEVMFRTP---- 1637 RT 389 GNYVHLVUDENFDPLDQWYSLILFKLKGUVGKVLMSVQGSKRKRKURVHLCTNTDPRYKEG 448 RT 1687 :|IT----WIDKXQAOVLVLSAQIAWSENVENTALSSMGGDA 1723 RT 486 --HGLLSK-SVQIANGLTIKMVDQDTPLKIMERL 516 RT 1724 APLHSVLSNVTEVTLVLAQSVLMEQ--PPLRRKL 1756 RT

RESULT 14

DYHC RAT ID - DYHC RAT STANDARD: PRT; 4644 AA.

AC P38550; Q63178; DT 01-FEB-1995 (Rel. 31, Created) DT 01-FEB-1995 (Rel. 31, Last sequence update) DT 15-SEP-2003 (Rel. 42, Last annotation update) DB Dynein heavy chain, cytosolic (DYHC) (Cyttoplasmic dynein heavy chain, cytosolic (DYHC)) (MAP1C) DE (MAP1C) GN DNEC1 OR DNC1L OR DNEC1L OR MAP1C. OS Rattus norvegicus (Rat). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Buteraria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI\_TaxID:10116; RN [1] RT SEQUENCE FROM N.A. RC STRAIN=Wistar; TISSUE=Brain; MEDLINE=3337615; PubMed=7630137; RA Zhang Z.; Tanaka Y.; Nonaka S.; Aizawa H.; Nakata T.; RA Hirokawa N. "The primary structure of rat brain (cytoplasmic) dynein heavy chain, a cytoplasmic motor enzyme." Proc. Natl. Acad. Sci. U.S.A. 90: 7928-7932 (1993). RN [2] RP SEQUENCE FROM N.A. RC STRAIN=Sprague-Dawley; TISSUE=Brain; MEDLINE=33264075; PubMed=7634232; RA Mikami A., Paschal B.M., Mazundar M., Vallee R.B.,

RT "Molecular cloning of the retrograde transport motor cytoplasmic  
 RT dynein (MAP 1C)." ;  
 RL "Neuro '90:797-796(1993)" ;  
 CC "FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular  
 CC retrograde motility of vesicles and organelles along microtubules.  
 CC Dynein has ATPase activity; the force-producing power stroke is  
 CC thought to occur on release of ADP.  
 -1- SUBUNIT: Consists of at least two heavy chains and a number of  
 CC intermediate and light chains.  
 -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC dynein heavy chains probably consist of an N-terminal stem  
 CC (which binds cargo and interacts with other dynein components),  
 CC and the head or motor domain. The motor contains six tandemly  
 CC linked AAA domains in the head, which form a ring. A stalk-like  
 CC structure (formed by two of the coiled coil domains) protrudes  
 CC between AAA 4 and AAA 5 and terminates in a microtubule-binding  
 CC site. A seventh domain may also contribute to this ring; it is not  
 CC clear whether the N-terminus or the C-terminus forms this extra  
 CC domain. There are four well-conserved and two non-conserved ATPase  
 CC sites, one per AAA domain. Probably only one of these (within AAA  
 CC 1) actually hydrolyzes ATP; the others may serve a regulatory  
 CC function.  
 -1- SIMILARITY: Belongs to the dynein heavy chain family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See, <http://www.ebi.ac.uk/announce/>)  
 CC or send an email to license@ebi.ac.uk.  
 CC  
 DR EML: D13896; BAA02996.; 1-  
 DR EMBL: L00505; AAA41103.; 1-  
 DR PIR: A38905; A38905.  
 DR InterPro: IPR04273; Dynein, heavy.  
 DR InterPro: IPR03593; AAA, ATPase.  
 DR Pfam: PF00308; AAA, 7.  
 KW Motor protein; dynein; microtubules; ATP-binding; Repeat; Coiled coil.  
 FT DOMAIN 1 1865 STEM (BY SIMILARITY).  
 FT DOMAIN 1666 2097 AAA 1 (BY SIMILARITY).  
 FT DOMAIN 2178 2450 AAA 2 (BY SIMILARITY).  
 FT DOMAIN 2254 2803 AAA 3 (BY SIMILARITY).  
 FT DOMAIN 2997 3165 AAA 4 (BY SIMILARITY).  
 FT DOMAIN 3187 3498 STALK (BY SIMILARITY).  
 FT DOMAIN 3551 3780 AAA 5 (BY SIMILARITY).  
 FT DOMAIN 403 4219 AAA 6 (BY SIMILARITY).  
 FT DOMAIN 48 69 COILED COIL (POTENTIAL).  
 FT DOMAIN 179 200 COILED COIL (POTENTIAL).  
 FT DOMAIN 453 476 COILED COIL (POTENTIAL).  
 FT DOMAIN 541 564 COILED COIL (POTENTIAL).  
 FT DOMAIN 1169 1201 COILED COIL (POTENTIAL).  
 FT DOMAIN 1229 1250 COILED COIL (POTENTIAL).  
 FT DOMAIN 1355 1371 COILED COIL (POTENTIAL).  
 FT DOMAIN 3187 3273 COILED COIL (POTENTIAL).  
 FT DOMAIN 3394 3498 COILED COIL (POTENTIAL).  
 FT DOMAIN 3135 3798 COILED COIL (POTENTIAL).  
 FT NP BIND 1904 1911 ATP (POTENTIAL).  
 FT NP BIND 2222 2229 ATP (POTENTIAL).  
 FT NP BIND 2293 2600 ATP (POTENTIAL).  
 FT NP BIND 2335 2942 ATP (POTENTIAL).  
 FT CONFLICT 1025 SR -> MP (IN REF. 2).  
 FT CONFLICT 1172 N -> D (IN REF. 2).  
 FT CONFLICT 2098 P -> A (IN REF. 2).  
 FT CONFLICT 2139 P -> V (IN REF. 2).  
 FT CONFLICT 2175 D -> A (IN REF. 2).  
 FT CONFLICT 2185 K -> Q (IN REF. 2).  
 FT CONFLICT 2366 L -> V (IN REF. 2).  
 FT CONFLICT 2382 T -> S (IN REF. 2).  
 FT CONFLICT 2463 G -> A (IN REF. 2).  
 FT CONFLICT 3129 A -> D (IN REF. 2).  
 FT CONFLICT 4131 R -> K (IN REF. 2).  
 R -> K (IN REF. 2).

PT	CONFLICT	4366	4366	F -> S (IN REF. 2).
PT	CONFLICT	4511	4511	A -> G (IN REF. 2).
SEQUENCE	4644 AA;	532240 MW;	8C6ABDBDF875D82	CRC64;
Query Match		3.54;	Score 99;	DB 1; Length 4644;
Best Local Similarity		19.54;	Pred. No. 28402;	Matches 109; Conservative 85; Mismatches 188; Indels 178; Gaps 25;
QY		41 LDPTQEPHLIISPSFLSUTIDANLATDPRFLILLGSPKLRLARGESPAYRPGGKTD	100	Db 1341 IDQMKEQPWSVQPRKURQNLQGLNLKNP --- PARLROYASVEFVQRLKGYMKIN 1395
QY		101 FL-FDPKKESTFEERSWYQWQNO-----DICKYGSIPPPDV-----	136	Db 1396 MLYVBLKSEA-LKDR-HWKQMLMGRHNWVNTSBLTQGQIWVUDLQRNNEAVKDVLLVAQ 1453
QY		137 ---EE---KLRLEWYQEQQLREHYOKKFQNSTYRSSVUDVLYFANCSDLIGFLN 189	1454 EMALBEPLKIQREWVNTYBLDLY-NYQKCR-----LIRGD 1489	
QY		190 ALLRTADLQWNSNAQLLQDCKSSKGNI-----SWBLGPNPSFLKKADIFIN-----	238	Db 1490 DLFLNKVKERHINSYSAMKLSPY --- YKVEREDALSNE-DKLNRTMADFWDIVQRRW 1542
QY		239 -----GS-----OLQEDBYQOLHKLKLKSTPKNAKLYGPGDVGQPRKT	275	Db 1543 VYLEGIFTGQSDAKHLIPVETORFQOSISTEFALMKKVSPLVMDVNLNQSVQSLRL 1602
QY		276 AKOMLKSFLKAGKAVI---DSVWMMHYHYNGRATRDEFNPDPDIF-----ISSVQYFO	328	Db 1603 ADLIGKTOKALGAYLERBRSSPFRYVG-----DEDLIBLIGNSKVNQVAKLQHF- 1652
QY		329 WVESTRPSKQVWGETSSAGGGAPLSDTFRAGFMWLDKLGSLARGMIEVVMRQVFFGA	388	Db 1653 -----KKMFGAVSSILNEDSSVV-----LGISSGREEEVNFKTP----- 1687
QY		389 GNYHLVDENFDPLADYWMISLLPKLVLGKTVLMAVSQGSKRKLRKVYHCTNTDNPRV--	445	Db 1688 -----VS1TEHPRKINEWLTVKEM--RTVTLAKLAESVTPVBFKATSDPNTVW 1739
Db		446 ---KEGDLTLYAINL---HNWTKYLRUPPYPSKPNQVKYLRPLGPGHGLSKSVQNLGT	499	Db 1740 IDKXQAOQVLVLSAQIANSVNEVNLNSVGGGN-----VGPLQSVISNVB--VT 1786
QY		500 LKAVDDQTL--PPLMEKPL 516		Db 1787 LNTYLAQSTVMEQPPRRKRL 1806
RESULT	15			
TMH_HARIN	ID TMH_HARIN	STANDARD;	PRT;	629 AA.
AC	PT1366;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DB	Putative type III restriction modification system HindIVP enzyme mod			
DB	(BC 2.1.1.72) (HindIVP-methyltransferase) (M.HindIVP).			
GN	HN1056			
OS	Haemophilus influenzae			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=rd / KM20 / ARCC 51907;			
RX	Medline=95350630; Pubmed=7542800;			
RA	Pleisemann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F.,			
RA	Reich, J.-P., Boughey, B.A., Merrick, J.M.,			
RA	Karlavage, A.R., Sutton, G., FitzHugh, W., Fields, C.A., Gocayne, J.D.,			
RA	McKenney, K., Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M.,			
RA	Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M.,			
RA	Weldman, J.P., Phillips, C.A., Springer, T., Hedstrom, B., Cotton, M.D.,			
RA	Uterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C.,			

Search completed: October 22, 2003; 20:24:56  
Job time : 29 SECs

Job time : 29 sec